Design principles and control mechanisms of signal transduction networks

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von

Herr Dipl.-Phys.BerndBinder

geboren am 06.08.1970 in Ilshofen

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Thomas Buckhout, PhD

Gutachter:

1. Prof. Dr. Reinhart Heinrich
2. Prof. Dr. Hans-Peter Herzel
3. Prof. Dr. Thomas Höfer

eingereicht am:	31. Januar 2005
Tag der mündlichen Prüfung:	31. März 2005

Ignorance more frequently begets confidence than does knowledge: it is those who know little, not those who know much, who so positively assert that this or that problem will never be solved by science.

Charles Darwin, The Descent of Man, 1871

Abstract

This work is based on the hypothesis that signal transduction networks in living cells are the result of an evolutionary development which is governed by mutation mechanisms and natural selection principles. This concerns their structural design as well as kinetic parameters. Therefore, it can be assumed that these properties have adopted values which imply certain optimal features with respect to the biological function of signal transduction.

Based on this working hypothesis, two approaches are presented to investigate the structural design and control mechanisms of signal transduction networks. Both strategies have as a common research objective the explanation of these properties of signalling networks using certain efficiency criteria.

In the first approach, covered in chapter 2, a model is developed to analyse the structural design of signalling networks. A simplified model is used to describe signalling systems consisting of receptors, kinases and phosphatases. This description includes important systems like MAP kinase pathways, the PI3K pathway, but also larger networks which exhibit complex crosstalk activations. Following the hypothesis mentioned above, optimisation principles have been applied to determine optimal network structures regarding certain biological functions. Two dynamic features of signalling systems are assumed to be crucial for the organism's surviving capacity: (i) To avoid autoactivation, e.g. due to stochastic fluctuations of receptor ligand or kinase concentrations, the signal off-state must be dynamically stable. (ii) The signal output should be of considerable magnitude, therefore the system should rather amplify than dampen the signal. To characterise network structures fulfilling both criteria simultaneously, a systematic analysis is performed for small networks consisting of up to seven kinases. The investigations reveal that for such networks the following two design principles hold true: (i) With increasing network size the connectivity decreases connoting an increasing specificity of kinase activities. (ii) The number of feedback cycles decrease with increasing network size indicating a decreasing tendency of downstream kinases to activate upstream kinases.

Based on the effect of small structural perturbations on the dynamic functions, a quantitative definition of the robustness of signalling networks is provided. Ratios of phosphatase and kinase activities maximising this robustness are identified.

The general validity of these design principles is supported by the analysis of a large kinase network containing 86 kinases which has been retrieved from the TRANSPATH database. This network is indeed characterised by a very low connectivity and a lack of positive feedback cycles.

This large signalling network is further investigated regarding several design properties. It is shown that the kinases group into two functional classes. About half of them can be regarded as "distributors" which means that they possess significantly more downstream targets than upstream activators. The other kinases act as "receivers", i.e. they possess more upstream activators than downstream targets. One identifies 25 input kinases which are not activated by other kinases and 27 output kinases without downstream targets.

All possible pathways from every input to every output kinase are calculated and the distribution of *shortest* pathway lengths is obtained. Interestingly, the most frequent shortest path length amounts to three corresponding to the length of MAPK pathways. Including also G-proteins leads to a signalling network of 94 components in which the most frequent shortest path length amounts to four, again corresponding to the length of MAPK pathways whose first kinase is activated by a G-protein.

When comparing the design and dynamic features of the TRANSPATH network to random networks, significant differences are observed: (i) Most random networks have feedback cycles. (ii) The total number of pathways tends to be larger in random networks. (iii) The fraction of shortest pathways is much higher in the TRANSPATH network. (iv) In the TRANSPATH network the relative frequency of pathway pairs without crosstalk is significantly higher. (v) The amplitudes and the signal durations of the kinase activation profiles tend to be less correlated in the TRANSPATH network than in random networks.

In order to understand evolutionary principles leading to acyclic networks, the procedure to generate random networks is altered by avoiding small cycles. Surprisingly, this additional constraint significantly increases the probability to generate cycle free networks.

In the second approach, described in chapter 3, control mechanisms of different signal transduction networks are analysed by applying Metabolic Control Analysis to transient activation profiles. Control coefficients on the signal amplitude, signal duration and integrated response are calculated.

First, the control mechanisms of a simple cascade model are compared to inhibitor experiments. The results confirm the predictions formulated by Heinrich et al. [2002] that kinases exert higher control on amplitude than on duration of the signal output,

whereas phosphatases have stronger control on duration.

Second, the control coefficients of activation profiles of ERK and other important components of a large EGF signalling model proposed by Schoeberl et al. [2002] are analysed. It can be observed that: (i) Only few reactions have considerably high control coefficients for the three characteristics amplitude, duration and integrated response. (ii) Kinase reactions generally have positive control on these features, while phosphatases tend to have a negative control. (iii) There is a high correlation between the control of duration and integrated response of the signal output (ERK), whereas the control coefficients of amplitude and duration are not correlated, i.e. these two features are independently controlled. (iv) The largest amplitude control coefficients are found for the MEK kinase indicating the highest average sensitivity to variations of reaction rates.

A more fundamental question regarding the design of signalling cascades is addressed in section 3.4. The optimal length of cascades which maximises the average sensitivity of the signal amplitude is detected. Whereas for pathways with mono phosphorylation steps the average sensitivity is almost independent of pathway length, cascades with dual phosphorylation steps show a maximal sensitivity for a pathway length of three or four kinases.

The presented results show that the structural properties of signalling networks drastically differ from those of random networks. Moreover, import characteristic features, such as the length of signalling cascades can be explained by optimisation principles.

Zusammenfassung

Diese Arbeit basiert auf der grundlegenden Annahme, dass Signaltransduktionsnetzwerke in lebenden Organismen als das Ergebnis eines evolutionären Prozesses betrachtet werden können, der sich durch Mutations- und Selektionsprinzipien auszeichnet. Dies betrifft sowohl das strukturelle Design als auch die kinetischen Parameter. Des-halb kann angenommen werden, dass diese Eigenschaften so beschaffen sind, dass sie bestimmte optimale Merkmale hinsichtlich der biologischen Funktion von Signaltransduktionsnetzwerken mit sich bringen.

Basierend auf dieser Hypothese werden zwei Ansätze vorgestellt, um strukturelles Design und Kontrollmechanismen von Signaltransduktionsnetzwerken zu untersuchen. Beide Strategien haben als gemeinsames Forschungsziel, Eigenschaften von Signaltransduktionsnetzwerken zu erklären, indem bestimmte Kriterien hinsichtlich ihrer Funktionsfähigkeit angelegt werden.

Beim ersten Ansatz in Kapitel 2 wird ein Modell entwickelt, welches das strukturelle Design von Signalnetzwerken analysiert. Ein vereinfachtes Modell wird benutzt, um Signalsysteme, bestehend aus Rezeptoren, Kinasen und Phosphatasen, zu beschreiben. Diese Beschreibung schließt Systeme wie MAP Kinase Wege, den PI3K Weg, aber auch größere Netzwerke, die komplizierte "Crosstalk"-Aktivierungen aufweisen, ein.

Es wird angenommen, dass zwei dynamische Eigenschaften dieser Systeme für die Überlebensfähigkeit der Organismen eine besonders wichtige Rolle spielen: (i) Um eine Autoaktivierung zu verhindern, z.B. gegenüber stochastischen Schwankungen von Rezeptorliganden oder Kinasekonzentrationen, muss der Signal-Aus Zustand dynamisch stabil sein. (ii) Das Ausgangssignal sollte eine beträchtliche Stärke aufweisen, weshalb das Signal eher verstärkt als gedämpft werden sollte. Um Netzwerk-Strukturen zu charakterisieren, die beide Kriterien gleichzeitig erfüllen, wird eine systematische Analyse von kleinen Netzwerken, die bis zu sieben Kinasen umfassen, durchgeführt. Die Untersuchungen machen deutlich, dass für solche Netzwerke die folgenden zwei Design-Eigenschaften gelten: (i) Mit steigender Netzwerkgröße verringert sich die Konnektivität, was einer steigenden Spezifität der Kinasen gleichkommt. (ii) Die Anzahl der "Feedback"-Zyklen nimmt mit zunehmender Netzwerkgröße ab, was eine abnehmende Tendenz anzeigt, dass "nachgeschaltete" Kinasen "vorgelagerte" Kinasen aktivieren.

Basierend auf den Auswirkungen von kleinen, strukturellen Störungen auf das dynamische Verhalten wird eine quantitative Definition von Robustheit von Signalnetzwerken angegeben.

Die allgemeine Gültigkeit dieser Design-Eigenschaften wird durch die Untersuchung eines großen Kinase-Netzwerks gestützt, das aus der Datenbank TRANSPATH entnommen wurde und 86 Kinasen umfasst. Das Netzwerk ist tatsächlich durch eine sehr niedrige Konnektivität und durch das Fehlen von positiven "feedback"-Zyklen gekennzeichnet.

Dieses große Signalnetzwerk wird hinsichtlich verschiedener Strukturmerkmale weitergehend untersucht. Es wird gezeigt, dass sich die Kinasen in zwei funktionelle Gruppen einteilen lassen. Ungefähr die Hälfte von ihnen kann als "Signalverteiler" betrachtet werden, d.h. sie besitzen bedeutend mehr nachgeordnete Zielkinasen als vorgeschaltete Kinasen. Die anderen Kinasen verhalten sich als "Signalempfänger", d.h. sie haben mehr vorgeschaltete, aktivierende Kinasen im Vergleich zu Zielkinasen. Man identifiziert 25 Eingangs-Kinasen, die nicht von anderen Kinasen aktiviert werden und 27 Ausgangs-Kinasen, die keine Zielkinasen aktivieren.

Alle möglichen Signalwege von jeder Eingangs-Kinase zu jeder Ausgangs-Kinase werden berechnet; anschließend wird die Verteilung der *kürzesten* Weglängen erstellt. Interessanterweise ist die am häufigsten auftretende Weglänge drei, was der Länge von MAP Kinase Wegen enstpricht. Wenn auch G-Proteine einbezogen werden, ensteht ein Netzwerk aus 94 Komponenten, in welchem die häufigste, kürzeste Weglänge vier ist. Dies stimmt ebenso mit MAP Kinase Wegen überein, deren erste Kinase von einem G-Protein aktiviert wird.

Wenn man das Design und die Dynamik des TRANSPATH Netzwerks mit den Eigenschaften von Zufallsnetzen vergleicht, zeigen sich aussagekräftige Unterschiede: (i) Die meisten Zufallsnetze besitzen "feedback"-Zyklen. (ii) Die gesamte Anzahl von Signalwegen ist typischerweise höher in Zufallsnetzen. (iii) Der relative Anteil von kürzesten Wegen ist im TRANSPATH Netzwerk viel höher. (iv) Im TRANSPATH Netzwerk ist die relative Häufigkeit von Paaren von Signalwegen, die keinen "Crosstalk" aufweisen, wesentlich höher. (v) Die Amplituden und Signaldauern der Aktivierungsprofile von Kinasen im TRANSPATH Netzwerk neigen dazu, weniger stark korreliert zu sein als in Zufallsnetzen.

Um evolutionäre Gesetzmäßigkeiten, die zu zyklenfreien Netzwerken führen, besser verstehen zu können, wird die Methode zur Erzeugung von Zufallsnetzen so geändert, damit insbesondere die Enstehung von *kleinen* Zyklen verhindert wird. Erstaunlicherweise erhöht diese zusätzliche Nebenbedingung die Wahrscheinlichkeit, Netzwerke ohne Zyklen zu generieren, beträchlich.

In einem zweiten Ansatz, der in Kapitel 3 beschrieben wird, werden Kontrollmechanismen unterschiedlicher Signaltransduktionsnetzwerke analysiert, indem die Metabolische Kontroll Theorie auf transiente Aktivierungsprofile angewendet wird. Es werden Kontrollkoeffizienten für die Signalamplitude, Signaldauer und das Integrierte Signal berechnet.

Zunächst werden die Kontrollmechanismen eines einfachen Kaskaden-Modells mit Inhibitor-Experimenten verglichen. Die Resultate bestätigen die Vorhersagen (Heinrich et al. [2002]), dass Kinasen eine höhere Kontrolle auf die Amplitude als auf die Signaldauer des Signalausgangs ausüben, und dass Phosphatasen in stärkerem Maße die Signaldauer kontrollieren.

Zweitens werden Kontrollkoeffizienten von Aktivierungsprofilen der ERK Kinase und anderen wichtigen Komponenten eines großen EGF Signal Modells analysiert, das von Schoeberl et al. [2002] vorgeschlagen wird. Es zeigt sich: (i) Nur wenige Reaktionen haben nennenswert große Kontrollkoeffizienten für die drei Kenngrößen Amplitude, Signaldauer und Integriertes Signal. (ii) Kinasereaktionen üben im Allgemeinen eine positive Kontrolle auf diese Größen aus, während Phosphatasen diese negativ kontrollieren. (iii) Es gibt eine hohe Korrelation zwischen der Kontrolle der Signaldauer und des Integrierten Signals des Signalausgangs (ERK), während die Kontroll-koeffizienten der Amplitude und der Signaldauer nicht korrelieren, d.h. diese beiden Merkmale werden unabhängig voneinander kontrolliert. (iv) Die größten Amplituden-Kontrollkoeffizienten werden für die MEK Kinase gefunden, was die höchste, durchschnittliche Sensitivität hinsichtlich der Variation von Reaktionen anzeigt.

Eine grundsätzlichere Frage zum Design von Signalkaskaden wird in Abschnitt 3.4 behandelt. Die optimale Länge von Kaskaden, welche die durchschnittliche Sensitivität der Signalamplitude maximiert, wird ermittelt. Während für Signalkaskaden mit einfachen Phosphorylierungs-Schritten diese Sensitivität fast unabhängig von der Kaskaden-Länge ist, zeigen Kaskaden mit zweifacher Phosphorylierung eine maximale Sensitivität für eine Kaskaden-Länge von drei oder vier Kinasen.

Die dargestellten Resultate zeigen, dass sich die strukturellen Eigenschaften von Signaltransduktionsnetzwerken auf drastische Weise von denjenigen von Zufallsnetzwerken unterscheiden. Darüber hinaus können wichtige charakteristische Merkmale wie die Länge von Signalkaskaden, ausgehend von Optimierungsprinzipien, erklärt werden.

Contents

Chapter 1

Introduction

1.1 Historical overview

Compared to metabolic networks, which are among the most thoroughly studied biological systems in the field of mathematical modelling, modelling of signal transduction is yet in an early stage. Nevertheless in recent years, an enormous amount of information has been accumulated about the components of various signalling pathways and their interplay. Such information is increasingly accessible in specific databases, such as TRANSPATH (Schacherer et al. [2001]) and STKE (Gough [2002]).

The MAPK signalling pathway has been modelled in successively greater detail since the early 1990s, and now boasts a greater proliferation of models than almost any other signalling pathway. The MAPK cascade as we know it today is a three-tiered cascade of kinases: MAPKKK $\longrightarrow$ MAPKK $\longrightarrow$ MAPK. It took several years to resolve even this basic topology of the cascade. In 1990 it was still not clear whether the kinases in the cascade were sequentially or simultaneously activated (Ahn and Krebs [1990]). A two-tiered ordered arrangement of the kinases had been proposed in 1991 (Ahn et al. [1991]), and in the span of two years from 1991 to 1993, the three-tiered core structure of the MAPK cascade was identified (Kyriakis et al. [1992]). Further refinements to the topology were made by the discovery that the MAPK cascade may involve scaffold proteins (Choi et al. [1994]).

The first simulations of the MAPK cascade were carried out in 1996 (Huang and Ferrell [1996]) in a kinetic model of MAPK demonstrating ultrasensitivity. This model uses experimental data, but also relies on estimations of certain rate constants. In 1997, Ferrell and Bhatt revealed that MAPKK phosphorylates p42 MAPK by a two-collision distributive mechanism rather than a single-collision processive mechanism (Ferrell and

Bhatt [1997]). This experimental model provided a mechanistic basis for understanding of how the MAPK cascade can convert graded inputs into switch-like outputs. In the same year, Burack and Sturgill confirmed by experiments and kinetic analysis of available data that the mechanism of ERK2 (a specific MAPK) activation by MEK1 (a specific MAPKK) in vitro is indeed non-processive (Burack and Sturgill [1997]). In 1998, Ferrell and Machleder showed that the MAPK cascade is activated essentially in an all-or-none fashion during Xenopus oocyte maturation (Ferrell [1998]). This behavior was proposed to arise from two known properties of the oocyte's MAPK cascade, a positive feedback loop and the cascade's intrinsic ultrasensitivity.

Following these initial model approaches, in the next phase additional components of the MAPK cascade were analysed and its dynamics were explored in detail. In 1999, Bhalla and Iyengar used simulations based on kinetic data available to propose that the MAPK cascade might participate in a bistable feedback loop (Bhalla and Iyengar [1999]). This MAPK model was one of the earliest to incorporate EGF activation in its description. In the same year, Kholodenko modelled the EGF signal transduction from the receptor to the RasGTPase (Kholodenko et al. [1999]) and in 2000, he suggested an oscillatory mechanism in a MAPK cascade model (Kholodenko [2000]). Levchenko et al. [2000] explored effects of scaffolding on the MAPK cascade. His model proposed that scaffold proteins may biphasically affect the levels of MAPK signalling and thereby reduce its threshold properties. Brightman and Fell [2000] showed through quantitative modelling that feedback inhibition of the MAPK cascade determines the duration of cascade activation. A further study on the MAPK system dynamics by Asthagiri and Lauffenburger [2001] demonstrated that negative feedback could enhance an upstream signal in the MAPK cascade.

Apart from incorporating still further cell biological details such as receptor traffic and transcriptional control, in the most recent phase of modelling another trend became apparent: Recent studies have begun to utilise both experiments and simulations in an integrative fashion. Schoeberl et al. [2002] have modelled the effects on MAPK due to receptor internalization of EGF receptors. In 2002, Bhalla et al. combined experiments and modelling to support MAPK involvement in a bistable feedback loop (Bhalla et al. [2002]). This study also considered transcriptional activation of MKP-1 as an important component of the history dependence of the cellular response. Swain and Siggia [2002] modelled the multisite phosphorylation of MAPK and suggested that it acts to improve signalling specificity. In a theoretical study, Somsen et al. [2002] showed that scaffolding could induce selectivity in different MAPK modules even if they shared the same kinases at some levels in the cascade. Mathematical models of other specific signalling processes

have also been proposed recently. An example is a model for the Wnt-pathway which is essential for cell differentiation in early cell development (Lee et al. [2003]).

All these kinetic models describe the combined action of participating reactions, such as stimulation of receptors, phosphorylations and dephosphorylations, binding to adaptor proteins, scaffolds or transcription factors. Some models are confined to stationary states (Huang and Ferrell [1996]), whereas others explore dynamic features (Kholodenko et al. [1999]; Schoeberl et al. [2002]; Asthagiri and Lauffenburger [2001]). These models are very valuable for understanding the behaviour of specific signalling pathways for which experimental data is available. Apart from describing the dynamics of these systems, the models have already revealed interesting regulatory properties, for example switch-like response or bistability. However, since the topology, that is the wiring of the participating components of these signalling networks, has to be provided as input, all these models are not capable to explain the underlying structural design. A first general approach to investigate how certain structural features influence the functional properties was developed by Heinrich et al. [2002]. It is described how a limited number of structural and kinetic key parameters, such as cascade length, crosstalk activity and divers rate constants, affect the functional properties signal amplification, signalling time and duration as well as stability of the signal off-state.

1.2 Research objectives

Although modelling of specific signalling processes is becoming more advanced, there is, in this field, a lack of general theory which can provide a basis for deeper understanding of design principles in general and the relation between the structural and the functional properties of signalling pathways in particular.

For metabolic systems, first promising approaches to relate structural and functional characteristics have been developed (Meléndez-Hevia et al. [1997], Stephani et al. [1999]; Ebenhöh and Heinrich [2003]). However, a straight forward transfer of these methods to signalling networks proves difficult due to several crucial differences between these two types of systems. Whereas the function of metabolism is to chemically convert initial substrates into final products, thereby creating a mass flow, in signalling systems information is transmitted without a net conversion of the participating compounds. Moreover, in metabolism one encounters two different kinds of components: the biochemical reactants undergoing conversions and the enzymes catalysing these conversions. In sharp contrast to that, the participating components in signalling pathways exert a multitude of different functions, as receptors, kinases, phosphatases, G-proteins, exchange factors,

adaptors, scaffolds, transcription factors etc. Further, reactants and catalysts are not strictly distinguishable since many enzymes are found in both roles, for example kinases which may either catalyse phosphorylations or act as target substrates. In consideration of these basic differences a new strategy has to be applied to disentangle structural principles of signalling networks.

One objective of this work is to understand which structural design properties enable such networks to perform their biological function. Here, the term "structural design" refers to the following characteristics:

- length of a kinase cascade,
- the phosphorylation mechanisms: dual or mono phosphorylation steps,
- number of cross activations (crosstalk) between distinguishable pathways,
- total number of interactions (network connectivity) among all kinases in a network,
- number of feedback cycles which can be counted in a network,
- distribution of signal outputs and inputs which the set of kinases in a network possess.

To a large extent, the biological function is determined by the dynamic behaviour of a signalling system. One can characterise this behaviour by means of the following key values of transient activation profiles:

- dynamic stability of the signal-off state, defined as the state when there is no receptor activation,
- signal amplification,
- signal duration,
- integrated response of an activation profile.

Chapter 2 focuses on the interrelations between these two sets of structural and functional quantities. Among others, the following problems are addressed: (1) Are there any general rules how a pathway should be structurally designed to allow for signal amplification? (2) How do signalling networks consisting of hundreds of mutually interrelated kinases maintain a stable signal off-state, thus avoiding self-activation in the absence of receptor stimulation? (3) What are the requirements for a signalling network

to remain robust against changes in their interconnections resulting from mutations or the action of drugs? (4) Why are there only few positive feedback cycles found in signalling networks compared to negative feedback regulations?

A general treatment of these problems will certainly run into extreme difficulties. Not only is the number of interacting compounds generally very high, they also exert a multitude of different functions as mentioned above. In the present work, first steps are presented towards a theory which interrelates in a general way structural features of signalling pathways with their dynamic properties. For the sake of simplicity the analysis of this chapter is confined to networks consisting exclusively of kinases, phosphatases and receptors. Of particular interest are their amplification properties and their dynamic stability depending on (i) kinase and phosphatase activities and (ii) the wiring of the network.

The basic strategy is to describe a vast number of structurally different, alternative designs to identify those with optimal dynamic features and compare their design properties with those actually found in contemporary cells. This task is performed by taking into account existing information on kinase interactions as documented in the signal transduction database TRANSPATH[1]. It is further shown that the structural and dynamic properties of the networks retrieved from TRANSPATH significantly differ from those of random networks.

Besides the structural design, functional properties depend strongly on regulatory mechanisms. In chapter 3, the regulatory principles of various signalling networks are investigated by applying Metabolic Control Analysis (MCA). The main challenge in conveying MCA to signalling systems is that here the transient states are more important than steady state conditions.

As a starting point (section 3.2), experiments with kinase and phosphatase inhibitors are described to underpin previous theoretical predictions (Heinrich et al. [2002]) about regulation of signal amplitude and signal duration of kinase cascades. For a direct comparison with the results of these experiments, a simple cascade model is used to study competitive and non-competitive inhibitors in silico. In section 3.3, it is analysed how the control on amplitude and duration of the activation profiles of different signalling components (e.g. of G-protein Ras and the kinases Raf, MEK and ERK) is distributed within a large model of the EGF network published by Schoeberl et al. [2002].

In section 3.4, an optimal cascade length is identified by applying optimisation criteria. In metabolic systems similar problems are often tackled by maximising certain fluxes. As an objective function of signalling pathways, the maximisation of the over-

[1] http://www.transpath.de

all sensitivity – defined as the standard deviation of control coefficients – of the signal amplitude is proposed. This objective function can be interpreted as an average "controllability" of the signal output. This approach seeks to extend earlier investigations (Goldbeter and Koshland [1981]; Goldbeter and Koshland [1982]; Huang and Ferrell [1996]) which focus on zero-order ultrasensitivity as an important attribute of signalling cascades.

Of particular interest in the present approach is a comparison between cascades consisting of mono phosphorylation cycles with cascades consisting of dual phosphorylation cycles. One finds that the average amplitude sensitivity of cascades consisting of mono phosphorylation steps is independent of the cascade length. In sharp contrast to that, cascades with dual phosphorylations possess an optimal cascade length of three or four activation steps in a wide range of parameters.

Chapter 2

Structural design of signal transduction networks

In this chapter a strategy is presented for understanding the interrelations between the structural design and dynamic properties of signal transduction networks. The basic hypothesis says: Stability of the signal off-state as well as signal amplification were important targets at the evolutionary optimisation of signal transduction networks. Stability of the steady state, when there is no signal present at the receptor (signal-off state), is of importance for signalling networks if they do not want to run the risk of autoactivation. Morover, a signal should be amplifyied or at least not damped out on the way from the cell membrane to the its target site. In search of structural designs which provide both characteristics, full sets of in principle possible network designs with a specific number of kinases and phosphatases are considered. The set of possible topological designs with seven kinases contains already more than $8.8 \cdot 10^8$ networks and marks therefore the upper limit of computational feasibility. Each network design of a given set is analysed with respect to its dynamic characteristics and structural properties. In this theoretical approach it is shown that the stability is closely related to the number of kinases, the total number of kinase activations and the number of feedback cycles within the network. Signalling networks with a high number of kinases, high kinase connectivities and/or cycles tend to be more unstable and therefore run the risk to display autoactivation. In contrast to that increasing phosphatase activity stabilises the off-state.

By regarding the effect of structural changes of signalling networks on their dynamic properties one arrives at conclusions concerning the robustness against such mutations (Ebenhöh and Heinrich [2003]).

The methods presented here are based on nonlinear differential equations for the concentrations of the participating kinases and algorithms for generating and analysing directed graphs.

Finally, the theory is applied to a real network by taking into account existing information on kinase interactions as documented in databases for signal transduction. The design of a large kinase network retrieved from the database TRANSPATH shall be investigated. This analysis reveals that several structural characteristics required for pathway stability are indeed fulfilled for a kinase network of such large size.

The kinase network is then further investigated with respect to other significant design and functional properties. This involves the length of all distinguishable pathways therein, the distribution of signal inputs and outputs of the kinases, crosstalk strength and the correlation of amplitudes and signal durations of kinase activation profiles. To evaluate to what extent the values of these measures are significant for the kinase network under consideration, they are compared to properties of randomised networks of the same size (Milo et al. [2002]). It can be shown that there are significant differences in this regard.

2.1 The model

2.1.1 Basic model assumptions

Signalling networks which are analysed in this chapter consist of receptors, kinases and phosphatases and have the following properties: (1) each kinase may occur in an active and in an inactive form; (2) activation of kinases takes place by phosphorylation; (3) in their active forms kinases may phosphorylate other kinases; (4) inactivation of the kinases takes place by dephosphorylation, catalysed by phosphatases which are constitutively active, i.e. they are not considered to be variable in time. Effects of multiple phosphorylation and auto-phosphorylation of single kinases are ignored. So far, the actions of adaptor proteins or scaffolds are also not considered. Accordingly, it is assumed that initial activation of kinases occurs by its direct interaction with a stimulated receptor R, i.e. the receptor itself has a kinase activity.

Three examples of signalling networks are given in Figure 2.1.1 where the active and inactive forms of kinase i are denoted by X_i^* and X_i, respectively. Each dashed arrow indicates an activating effect of one kinase on another kinase. Arrows pointing from active forms to inactive forms of kinases denote the action of phosphatases. To allow for a large variety of pathway structures, the possibility that a given kinase may

be activated by more than one other kinase and that a single kinase is able to activate several other kinases (Bhalla and Iyengar [2001]) is considered. This gives rise to pathway structures, some of which feature different branches or feedback cycles. The first of the three examples, Network A, represents the most simple arrangement of a receptor and three pairs of kinases and phosphatases in form of a linear cascade resembling the classical MAPK pathway. The other two networks in Figure 2.1.1 represent more complex structures which are also found in real cells (Kyriakis and Avruch [2001]). In network B one of the kinases activates two kinases, and in network C one of the kinases receives two activating inputs. For the latter network these interconnections give rise to a feedback cycle. In the lower panel the same three networks are described in a more compact representation by directed graphs. The nodes represent kinases and the arrows their mutual activations. Whereas these graphs focus on the action of kinases, it should be kept in mind that to each node a reaction is attached representing a deactivation process by a phosphatase. Our dynamic analysis is based on the following

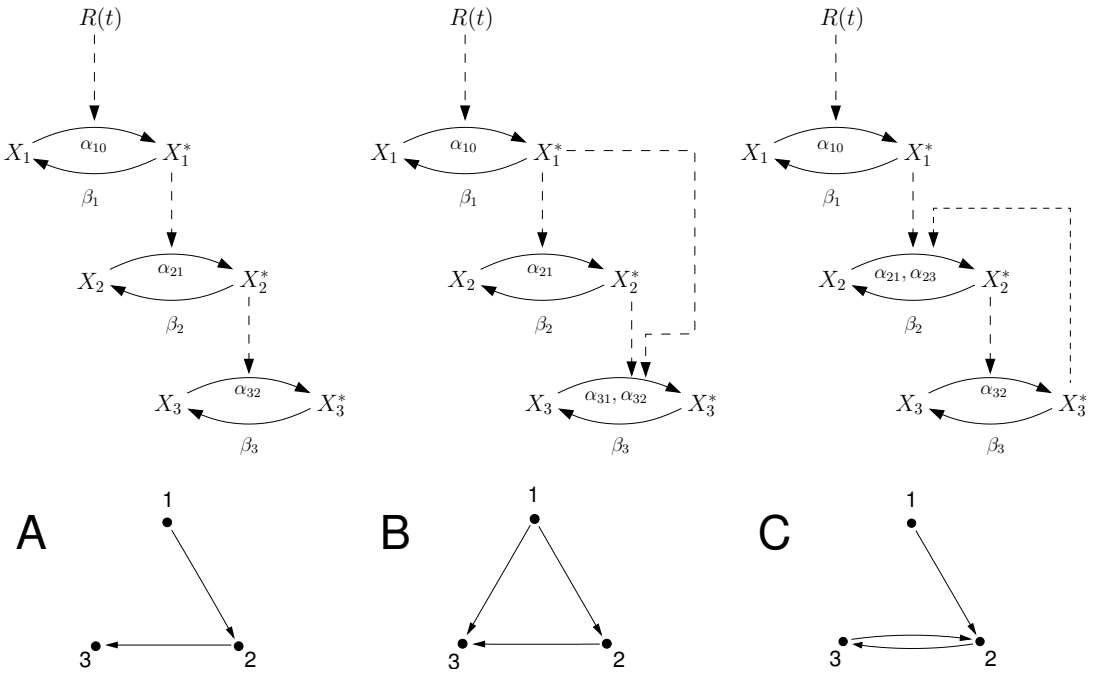


Figure 2.1: Different network structures with three kinases and three phosphatases. Upper panel: Interconversions of active and inactive forms of kinases catalysed by other kinases and by phosphatases. Lower panel: compact graph representation of the same networks.

set of differential equations (Heinrich et al. [2002])

$$\frac{dX_i^*}{dt} = \rho_i' R(t) X_i + \sum_{k \neq i}^{n} \alpha_{ik}' X_i X_k^* - \beta_i X_i^*. \tag{2.1}$$

The parameter n denotes the total number of kinases involved in a given pathway. Activation rates of kinases i by other kinases k are described by bimolecular terms $\alpha_{ik}' X_i X_k^*$ where the factors α_{ik}' are second order rate constants. The first order rate constants β_i characterise the dephosphorylation of the kinases by phosphatases. The first term in Eq. (2.1) describes kinase activation by the receptor. The corresponding rates are also written as bimolecular terms in which $R(t)$ denotes the time dependent concentration of the stimulated receptor and ρ_i' the corresponding second order rate constant. A transient stimulation of the receptor is described by an exponential decay, that is $R(t) = R \cdot \exp(-\lambda t)$, where λ characterises the half life of its active form. The case $\lambda \to 0$ results in $R = \text{const.}$ describing a sustained receptor activation. The concentrations of complexes resulting from protein–protein interactions are considered to be negligibly small, i.e. these complex concentrations do not appear as separate variables in the model. According to the reaction schemes in Figure 2.1.1 and the systems equations Eqs. (2.1), only the concentrations of active and inactive forms contribute to the total concentrations of kinases leading to conservation relations:

$$X_i + X_i^* = C_i = \text{const.} \tag{2.2}$$

This equation can be used to eliminate X_i from Eq. (2.1) resulting in

$$\frac{dX_i^*}{dt} = \left(\rho_i R(t) + \sum_{k \neq i}^{n} \alpha_{ik} X_k^* \right) \left(1 - \frac{X_i^*}{C_i} \right) - \beta_i X_i^*, \tag{2.3}$$

where $\rho_i = C_i \rho_i'$ and $\alpha_{ik} = C_i \alpha_{ik}'$ denote pseudo-first order rate constants.

2.1.2 Structural properties of signalling networks

As already shown in the lower panel of Figure 2.1.1 the signalling networks can be drawn as directed graphs. They consist of nodes representing the kinases, and of arrows representing activation of these kinases by other kinases. These graphs focus on the kinases whose interactions perform the signalling routes. However, the dynamic behaviour of the networks described by Eq. (2.3) also depends on the phosphatases acting on each kinase.

In the following, only those networks which cannot be decomposed into subnetworks of lower size shall be taken into account. In these network types, each node is connected to at least one other node (weakly connected graphs). Any of these graphs representing a special signalling network can be characterised by the following structural properties: the number n of nodes, the number e of edges for the activations, and the number c_L of cycles of length L. A cycle of length L in a directed graph is defined as a closed loop consisting of L edges, and no node may occur more than once within this loop. The outdegree and the indegree of a node are defined by the number of edges leaving this node or entering this node, respectively. The mean outdegree d of a network which equals its mean indegree and a related property, the network connectivity κ, are defined by

$$d = \frac{e}{n} \tag{2.4}$$

and

$$\kappa = \frac{e}{n(n-1)} = \frac{d}{n-1}, \tag{2.5}$$

respectively. These measures indicate how many activations are taking place among a given number of kinases. In biological terms κ characterises on average the unspecificity of kinase activations in a given network, and its value varies in the range of $1/n \leq \kappa \leq 1$. As $n(n-1)$ is the maximum number of activations within a network of n kinases $\kappa = 1$ holds for the fully connected network in which every kinase activates all other kinases (see graph No.1 in Figure 2.1.2). Similarly a scaled measure σ for the number of cycles in a certain network with n nodes can be introduced. The cycles are counted irrespective of their length L, and this number c is scaled by the maximum number of cycles $c^{\max}$ which can in principle occur in networks of size n. Thus, $c^{\max}$ is calculated as the total number of cycles in the fully connected graph of size n. The cycle density σ is defined as follows:

$$\sigma = \frac{c}{c^{\max}}, \quad \text{with} \quad c^{\max} = \sum_{i=2}^{n} \frac{n!}{i(n-i)!}, \tag{2.6}$$

and its value varies in the range of $0 \leq \sigma \leq 1$.

In Figure 2.1.2 all possible network designs of three kinases are depicted together with their structural properties. As previously stated, nodes represent active kinases and directed edges represent activations between kinases. In Table 2.1 the numbers of possible network designs with $n = 2 \ldots 7$ nodes are listed. The maximum number of feedback cycles for each set of n-networks is also given. The number of networks, which must be at least weakly connected, increases strongly with the number of kinases n and

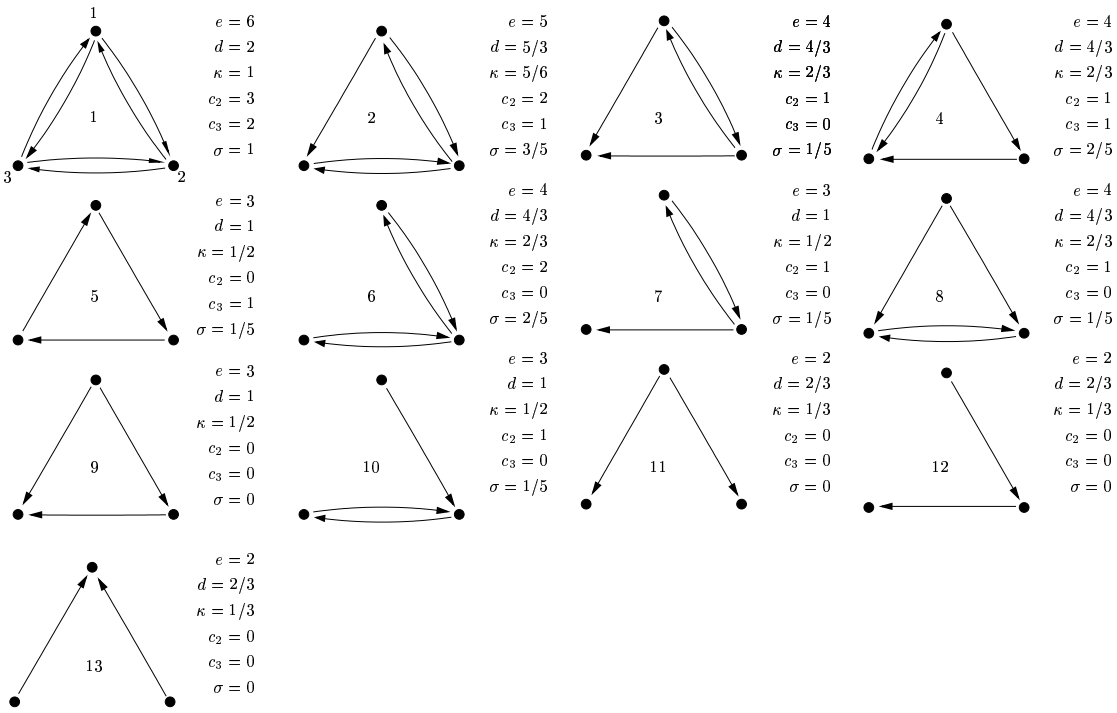


Figure 2.2: The full set of different network structures with three kinases ($n = 3$). The nodes depict the active kinases and the edges show the activations. Graphs No.12, 9, and 10 refer to the three examples considered in Figure 2.1.1.

can be found using an application of the Pólya enumeration theorem (Harary [1994]). These networks differ significantly from each other which means that no permutation of nodes can transfer one network into another. They are described as non-isomorphic to each other. For a more detailed list of numbers of network designs see Appendix A.2.

number of nodes	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
number of designs	2	13	199	9,364	1,530,843	880,471,142
max. number of cycles	1	5	20	84	409	2,365

Table 2.1: Numbers of network designs and maximum numbers of cycles depending on the number of nodes (kinases).

Concerning the primary activation of all kinases, one must accurately attach appropriate receptors to specific kinases in the network. Hence, a *minimum* set of receptors is

considered whose stimulation results in a subsequent activation of all kinases. For some networks the attachment of a receptor to a single input kinase may be sufficient. This is the case if the corresponding graph contains directed paths from this kinase to all other kinases. For other networks one might initially have to stimulate more than one kinase by receptors. This is particularly the case if a network contains several kinases without activations from other kinases. There are also networks where each kinase is equally suited for a primary activation. In this case the corresponding graphs are strongly connected and a receptor is attached to a kinase where the best amplification performance is guaranteed (for a definition of amplification see the next section 2.1.3. If a kinase i is activated by a receptor, then ρ_i' in Eqs. (2.1) differs from zero.

2.1.3 Dynamic properties of signalling networks

For the time independent states for the concentrations X_i^* only solutions of Eqs. (2.3) within the range $0 \leq X_i^* \leq C_i$ are of interest since otherwise the concentrations of active or inactive kinases would be negative. Positive steady state solutions for permanent receptor activation ($R \neq 0$) can generally not be written in an explicit form due to the nonlinear character of Eqs. (2.3). For $R = 0$, equation system (2.3) always has a time independent solution, $X_i^* = 0$, representing the signalling off-state. In this case, another stationary state with $X_i^* \neq 0$ resulting from autoactivation of the pathway may also exist. As an example, a network is considered which contains a cycle of length $L = 2$ between kinases 1 and 2 both of which activate a third kinase (see network 3 in Fig. 2.1.2). The steady state conditions read in this case

$$\dot{X}_1^* = \alpha_{12} X_2^* \left(1 - \frac{X_1^*}{C_1}\right) - \beta_1 X_1^* = 0 \tag{2.7}$$

$$\dot{X}_2^* = \alpha_{21} X_1^* \left(1 - \frac{X_2^*}{C_2}\right) - \beta_2 X_2^* = 0 \tag{2.8}$$

$$\dot{X}_3^* = \alpha_{32} X_2^* \left(1 - \frac{X_3^*}{C_3}\right) + \alpha_{31} X_1^* \left(1 - \frac{X_3^*}{C_3}\right) - \beta_3 X_3^* = 0. \tag{2.9}$$

Besides the off-state, $X_1^* = X_2^* = X_3^* = 0$, this equation system has the steady state solution

$$X_1^* = \frac{C_1 C_2 (\alpha_{12}\alpha_{21} - \beta_1\beta_2)}{\alpha_{21}(\alpha_{12} C_2 + \beta_1 C_1)},\ X_2^* = \frac{C_1 C_2 (\alpha_{12}\alpha_{21} - \beta_1\beta_2)}{\alpha_{12}(\alpha_{21} C_1 + \beta_2 C_2)}, \tag{2.10}$$

where the corresponding X_3^*-value is obtained by introducing the solutions for X_1^* and X_2^* into Eq. (2.9). The solutions for all X_i^* are proportional to the factor $\alpha_{12}\alpha_{21} - \beta_1\beta_2$.

Eq. (2.10) describes an autoactivated state which only exists when the phosphatase activities are low compared to the kinase activities, that is, for

$$\alpha_{12}\alpha_{21} > \beta_1\beta_2. \tag{2.11}$$

It can easily be proven that also for other networks autoactivated states may only occur if the corresponding graphs contain cycles (see Appendix A.3).

An intriguing question concerning the proper functioning of a signal transduction pathway is, if in the absence of receptor activation the off-state is dynamically stable. Stability means that the signalling network returns to this state after small perturbations such as those resulting from spurious activation of the kinases. In the case of instability, autoactivation may occur where the network will spontaneously leave the off-state, approaching a stationary state with non-zero concentrations for the active forms of some kinases, as in the state represented by Eq. (2.10). Stability analysis is performed by considering the spectrum of the eigenvalues of the Jacobian $\mathbf{J} = (J_{ik})$ of the differential equation system (2.3). For the off-state the Jacobian reads

$$J_{ik} = \alpha_{ik} - \beta_i\delta_{ik} \quad , \quad \begin{cases} \alpha_{ik} \neq 0 & \text{if kinase } i \text{ activates kinase } k \\ \alpha_{ik} = 0 & \text{otherwise} \end{cases} \tag{2.12}$$

δ_{ik} denotes the Kronecker symbol. Stability requires that the eigenvalues of this matrix have only negative real parts. For example, stability analysis of the signalling off-state for the network represented by graph 3 in Figure 2.1.2 reveals that the autoactivated state Eq. (2.10) is stable whereas the off-state is unstable. When condition (2.11) is not met, the autoactivated state does not exist and the off-state is stable. Concerning the stability of the off-state it is worth mentioning that the results presented in this work for mass action kinetics are also representive for less simplified kinetic rate equations, for example Michaelis-Menten or Hill equations. This is due to the fact that pathway stability against small perturbations is determined by the properties of the Jacobian of the off-state. This state becomes independent of saturation phenomena, i.e. also in more complex cases the Jacobian $\mathbf{J}$ for $X_i^* = 0$ is nevertheless described by Eq. (2.12).

During transduction of a signal from a stimulated receptor to a target kinase, amplification or dampening may occur. As examples the transient activation of the three pathways A, B, and C depicted in Figure 2.1.1 are considered. Receptor stimulation occurs for $t = 0$. Figure 2.1.3 shows the exponentially decreasing concentration of the active receptor for $t > 0$, and the three other curves are the time dependent activation curves of the third kinase for the respective pathway. For pathways A and B the off-states

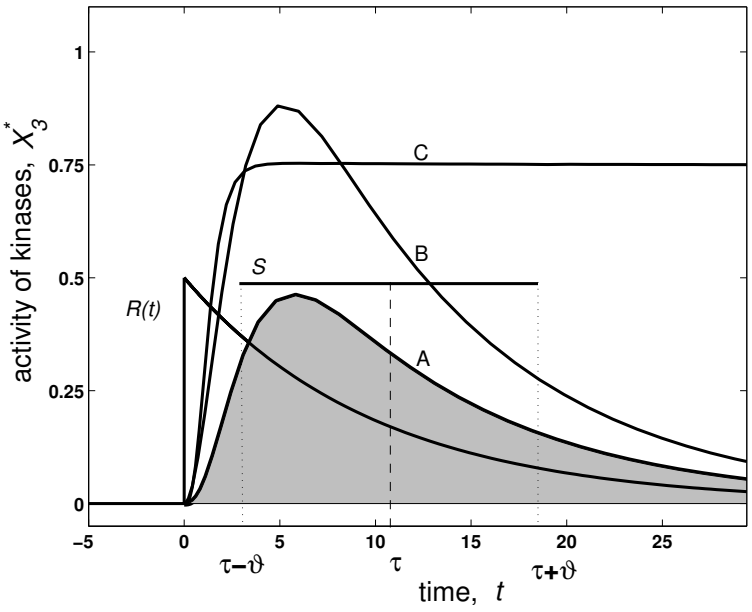


Figure 2.3: Time dependent changes of receptor $R(t)$ and kinase concentrations X_3^* in the networks A, B and C from Figure 2.1.1. In each network the receptor is attached to kinase 1 and has the same activation profile. Parameter values: $\alpha_{ik} = 1.0$, $\beta_i = 0.9$, $\lambda = 0.1$. Signalling amplitude S, duration ϑ, time τ and integrated response I (shaded area) are indicated for the output kinase of network A: $S_3 \approx 0.47$, $\vartheta_3 = 6.7$, $\tau_3 = 11.3$, $I_3 = 6.4$.

are stable. Accordingly, the target kinases are transiently activated and the networks return eventually to their off states. Compared to these responses, pathway C shows a completely different behaviour. The off-state is unstable and the network approaches a stable activated state which persists even after the decay of active receptor.

Different definitions for amplification or dampening may be envisaged. A convenient characterisation of amplification properties is based on a comparison between the concentrations of the active forms of kinases having different positions in the pathway or by comparing these concentrations to the concentration of the stimulated receptor. For complex signalling networks as depicted in Figures 2.1.1 or 2.1.2 it is convenient to define an overall amplification (or overall dampening) by comparing the highest values of the concentrations of the active kinases with the initial concentration of the receptor (R(0)). According to this definition pathway B amplifies the signal whereas in pathway A dampening occurs for the given parameter values. This concept of amplification is applicable to transient activation of pathways but also to stationary states which are

attained for permanent receptor stimulation ($\lambda \to 0$).

Amplification properties of the various networks can be analysed in principle by solving Eq. (2.3) numerically. However, determining the maxima of the concentrations of active kinases may render to be rather cumbersome. It is therefore more convenient to analyse amplification in terms of average activation amplitudes S_i. This idea refers to the concept of averaged signalling amplitudes introduced in Heinrich et al. [2002]. For a given kinase i such an amplitude is defined as follows:

$$S_i = \frac{I_i}{2\vartheta_i} \ , \tag{2.13}$$

where the "integrated response" I_i (Asthagiri and Lauffenburger [2001]) and ϑ_i are defined in the following way:

$$I_i = \int_0^\infty X_i^*(t)\ dt, \tag{2.14}$$

$$\vartheta_i^2 = \frac{1}{I_i}\int_0^\infty t^2 X_i^*(t)\ dt - \tau_i^2 \ \text{ with } \ \tau_i = \frac{1}{I_i}\int_0^\infty t X_i^*(t)\ dt. \tag{2.15}$$

The quantity I_i corresponds to the area under activation curve $X_i^*(t)$, ϑ_i characterises signal duration and τ_i denotes the average signal propagation time to the kinase i (Heinrich et al. [2002]). How these dynamic quantities are related to each other is depicted in Figure 2.1.3 for the activation profile of the third kinase of network A. The amplitude S is the height of a rectangle whose length is 2ϑ and whose area equals I (shaded). A signal is considered to be amplified in a network if a kinase i exists with $S_i \geq R(0)$. Also according to this defintion in pathway A dampening occurs because not only the maximum value of the activation profile is smaller than $R(0)$ but also S_3 (see legend to Figure 2.1.3). In that respect network B has an optimal structural design (for the given parameter values) because it possesses a stable off-state and additionally amplifies signals.

2.1.4 Dynamic properties of networks depending on their size, connectivity, number of cycles and kinetic properties

Concerning amplification, a proper choice has to be made for initial kinases, activated by the receptor, and for a final output kinase to which the signal is transmitted. Eventually that kinase is selected as the signal output for which a maximum amplification, i.e. a maximal average amplitude, is observed. As already mentioned, input kinases are those to which receptors have to be attached, in order to ensure that all kinases in the network

are activated after receptor stimulation. Note that stability of off-states is independent of the choice of input and output kinases and the strength of the receptor activation λ. To obtain explicit solutions and a qualitative description of the interrelation of structural design and dynamics properties, the kinases as well as the phosphatases are assumed to have identical kinetic properties as characterised by their rate constants: $\alpha_i = \alpha$ and $\beta_i = \beta$. This allows for dividing the equation system (2.3) by α which equals a time rescaling ($t \rightarrow t' = t\alpha$). In a more compact form Eq. 2.12 can be rewritten as

$$\mathbf{J} = \mathbf{A}^T - \frac{\beta}{\alpha}\mathbf{I}, \tag{2.16}$$

where $\mathbf{A}^T$ describes the transposed adjacency matrix of the underlying graph and $\mathbf{I}$ denotes the unity. To obtain also explicit solutions for the calculation of the signal amplitudes it is moreover assumed that the component kinases are in a low activated state (Heinrich et al. [2002]), i.e. they are phosphorylated to a low degree ($X_i^* \ll C_i$). This particular assumption leads to a linearisation of equation system (2.3) (see Appendix A.1). Due to this simplification the kinetics of signalling networks can be described with one parameter, β/α, which reflects the phosphatase activities normalised to the kinase activities.

First the stability of off-states and amplification properties of the family of 13 networks consisting of three kinases is analysed. Which kinases are chosen as signal input and signal output are listed in Table 2.2. The kinase numbers refer to the numbering of kinases in Fig. 2.1.2. In the following "stable network" simplifying denotes a network

network	1	2	3	4	5	6	7	8	9	10	11	12	13
input kinase	1	2	1	1	1	2	1	1	1	1	1	1	2 and 3
output kinase	3	2	3	3	1	2	1	3	3	2	3	3	1

Table 2.2: Choice of input and output kinases for 3-kinase networks. Input kinases are those for which a *minimum* set of receptors can be attached to get a fully activated network. If more possibilities are available those kinases are chosen as receptor inputs for which a maximum amplification is guaranteed. An output kinase is chosen under similar criteria: the output kinase exhibits the maximum amplification. The numbering refers to Figure 2.1.2.

possessing a stable off-state.

Figure 2.4 gives an overview of the stability and amplification properties of the family of 13 networks consisting of only three kinases. The ordinate represents the values of

the normalised rate constant β for phosphatases. On the abscissa of the diagram the networks are grouped according to the mean outdegree of the corresponding graphs. The blank part of any bar characterises that region of β-values for which the corresponding network is unstable. The shadowed part is the region where the corresponding network is stable and displays amplification. In all cases β-values above the shadowed parts of the bars result in stable networks with dampening properties. The results of this

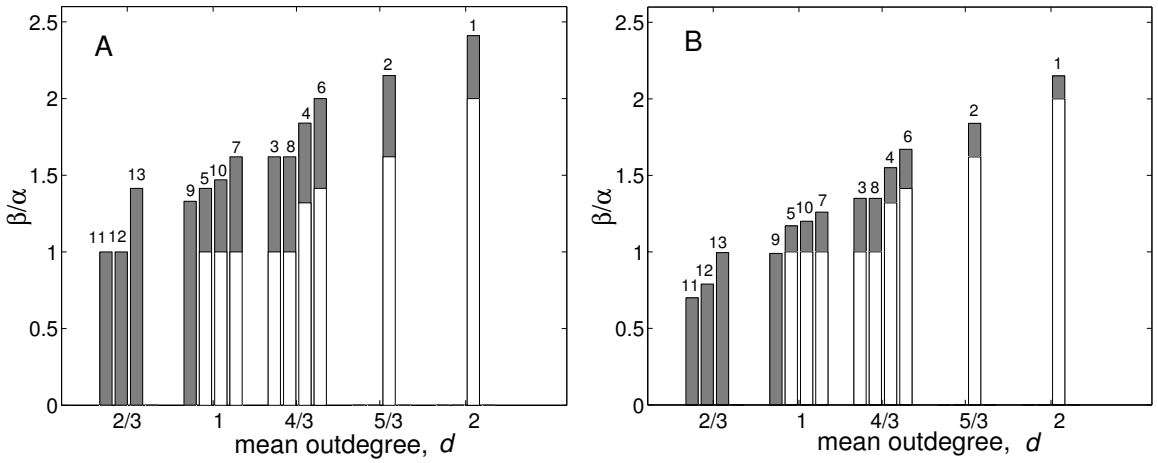


Figure 2.4: Stability and amplification of signalling networks with three kinases and three phosphatases for permanent receptor activation with $\lambda = 0$ (A) and for transient receptor activation with $\lambda = 0.1$ (B). Depending on the normalised activity of phosphatases and on the mean outdegree of the corresponding graphs one obtains regions of unstable off-states (blank parts of the bars) and of stable off-states together with signal amplification (shaded parts). For β/α-values above the bars the signal is dampened.

analysis can be summarised as follows:

1) The four graphs without any cycle (networks 9, 11 to 13) are always stable, but they show slight differences in their amplification properties. For example, network 13 displays an amplification over a larger range of phosphatase activities than the other three graphs. This results from the fact that the output kinase of network 13 receives activation from two kinases which themselves are activated by the receptor.

2) All networks with cycles have a region of β-values where the off-states are unstable. The networks with only one cycle (networks 3, 5, 7, 8, 10) have the same critical β-value below which instability occurs. With an increasing number of cycles the boundaries for instability are shifted toward higher β-values, with some slight

differences in the effects of cycles of length 2 or 3. Network 1 which is represented by a totally connected graph, has a maximum number of 5 cycles and is therefore characterised by the largest region of instability.

3) There is a general tendency that networks with higher numbers of interactions are less stable, as becomes evident by an increase of the instability region with an increasing mean outdegree.

4) Compared to the case of permanent receptor activation (Fig. 2.4A, $\lambda = 0$) transient receptor activation (Fig. 2.4B, $\lambda = 0.1$) results in narrowing of the region of β-values where amplification is observed. Since the critical β-values for instability of the off-states are independent on the receptor only upper boundaries of the amplification regions are affected when decreasing the life-time of the receptor.

Figure 2.5 shows regions of stability and regions of amplification for networks consisting of more than three pairs of kinases and phosphatases. The abscissa represents the

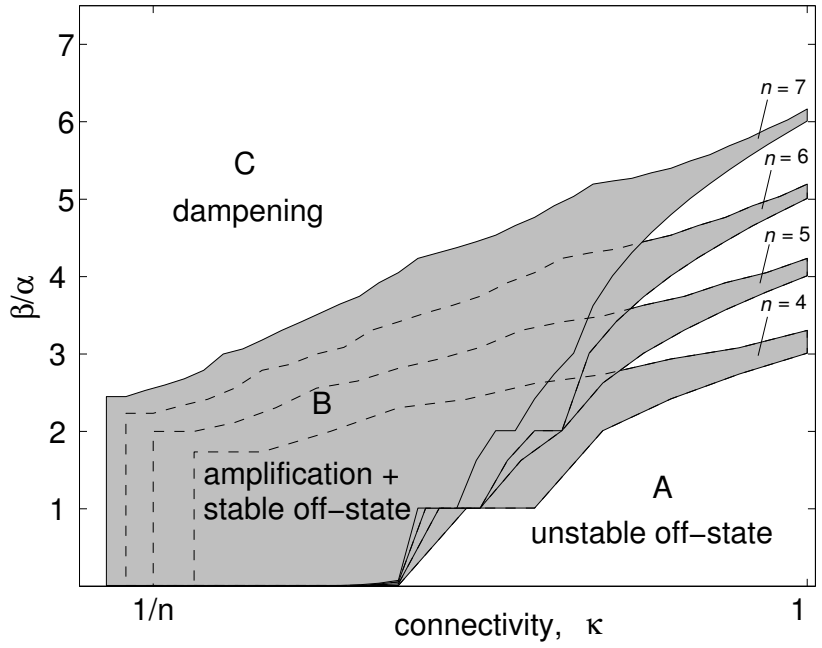


Figure 2.5: Effect of network connectivity κ and phosphatase activity β/α on stability and signal amplification for networks of different size. In the grey shaded parameter region networks exist which have a stable off-state and exhibit signal amplification.

network connectivity κ which, according to Eq. (2.5), is directly related to the mean outdegree d. Due to normalisation with respect to the number of kinases, κ is better suited

for comparing the properties of signalling networks of different size. Figure 2.5 shows that for each family of networks having the same size n there are three distinct regions **A**, **B**, and **C** in the $(\kappa, \beta/\alpha)$-plane, characterising parameter combinations leading to different dynamic properties of the networks. In region **A** all networks of a given family have an unstable signal off-state, in region **B** (shaded) stable networks exist showing amplification properties, and in region **C** all networks of the given family have a stable signal off-state and show a dampening in the signalling amplitudes. The borders between regions **A** and **B** and between **B** and **C** are shifted towards higher β-values when n increases, that is, networks with a higher number of kinases tend to be more unstable. Keeping in mind that biological cells contain several hundred different types of kinases, one may draw the conclusion that real kinase networks (see section 2.2) must exhibit a low connectivity to avoid autoactivation.

To illustrate the counteracting effect of network size n and phosphatase activity on stability, the distributions of maximum eigenvalues $\lambda_{\max}$ of Jacobians are considered. The histograms in Figure 2.6 show the distributions of the real parts of maximum

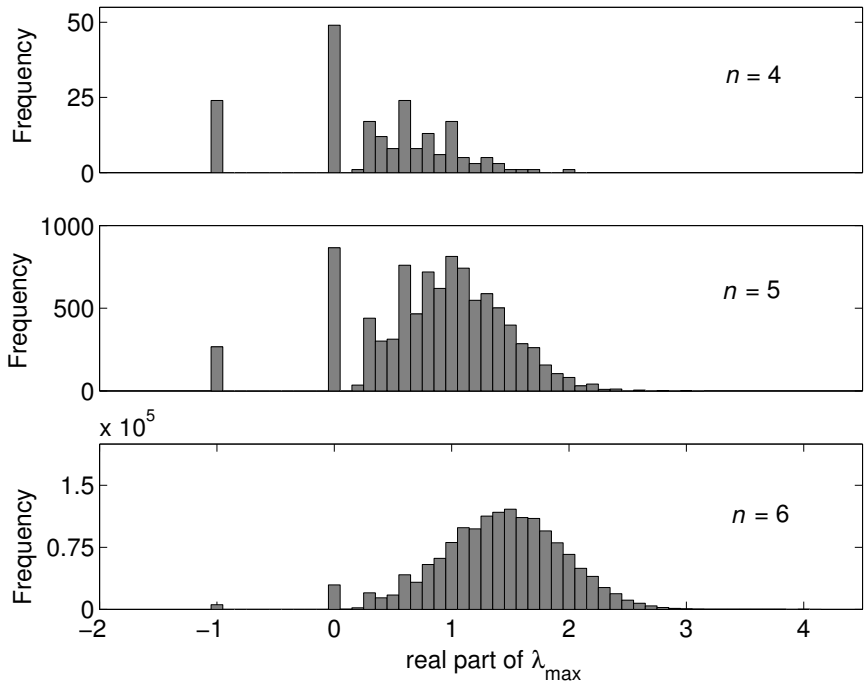


Figure 2.6: Distribution of largest eigenvalues $\lambda_{\max}$ of the Jacobians of signalling networks with 4, 5 and 6 kinases. Depicted are the real parts of $\lambda_{\max}$ (spectral radius). The rate constants of phosphatases are set to $\beta/\alpha = 1.0$.

eigenvalues for the full set of networks for three different system dimensions (n = 4, 5, 6). The rate constants of phosphatases are always set to $\beta/\alpha = 1.0$. The abscissa represents the real part of maximum eigenvalue and the ordinate the number of networks with a certain $\lambda_{\max}$-value within small intervals ($\Delta\text{Real}(\lambda_{\max}) = 0.1$). The percentage of stable networks ($\text{Real}(\lambda_{\max}) < 0$) decreases towards higher system dimension. For the given parameter values there are about 13%, 3% and 0.4% stable systems, respectively. Similarly the relative number of networks with $\text{Real}(\lambda_{\max}) = 0$ decreases with higher n. These particular networks contain exactly one cycle. On the other hand the percentage of systems with stable off-state becomes higher with increasing phosphatase activity. Increasing β/α results in a shift of the histograms to the left which becomes clear by the definition of J_{ik} (see Eq. (2.16)). In Figure 2.7 the dependence of the percentage of stable networks ($N_{st.}/(N_{st.} + N_{inst.})$) is depicted as a function of β. The plateaus for $0 < \beta/\alpha \leq 1$ are due to the fact that there are no networks for $-1 < \text{Real}(\lambda_{\max}) < 0$ (when $\beta/\alpha = 1$), which can be seen in Figure 2.6. Changing β/α from 1 to 0 keeps the number of stable networks constant.

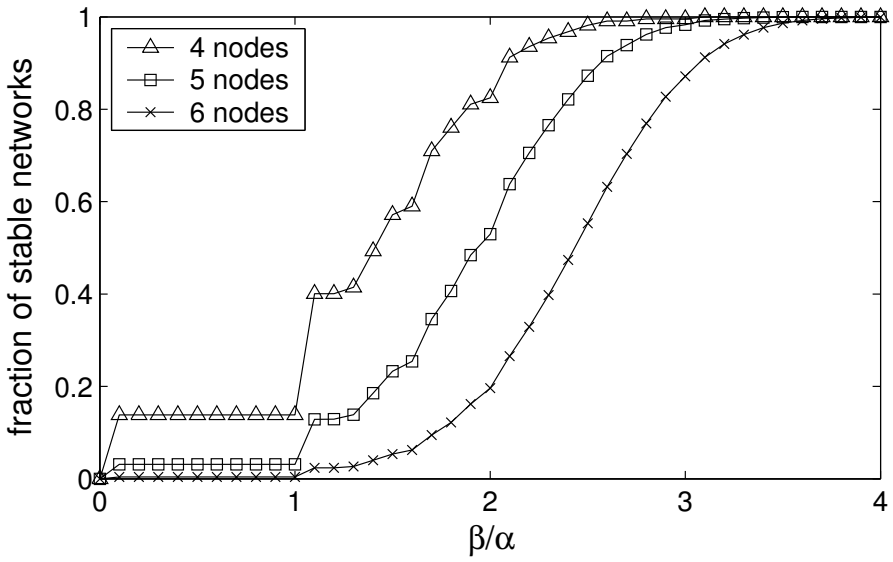


Figure 2.7: Dependence between the fraction of stable networks and phosphatase activity for networks of size n = 4, 5, 6.

The grey shaded $(\beta/\alpha, \kappa)$-parameter region **B** in Figure 2.5 contains signalling networks which are stable and amplifying at the same time. Therefore, this area can be considered as the biologically meaningful parameter region of signalling networks. Only within this region, networks exhibiting the two characteristics simultaneously exist. On

closer inspection one finds a more detailed partitioning. A distinction can be drawn between four different subregions $\mathbf{B_1}$, $\mathbf{B_2}$, $\mathbf{B_3}$, $\mathbf{B_4}$. They are distinct from each other with respect to different combinations of dynamic characteristics the networks exhibit therein. In Figure 2.8 the composition of the four subregions is illustrated by considering, as an example, the biological meaningful parameter region of networks of size $n = 6$. Markers on the outer lines indicate the discrete character of the κ-values existing

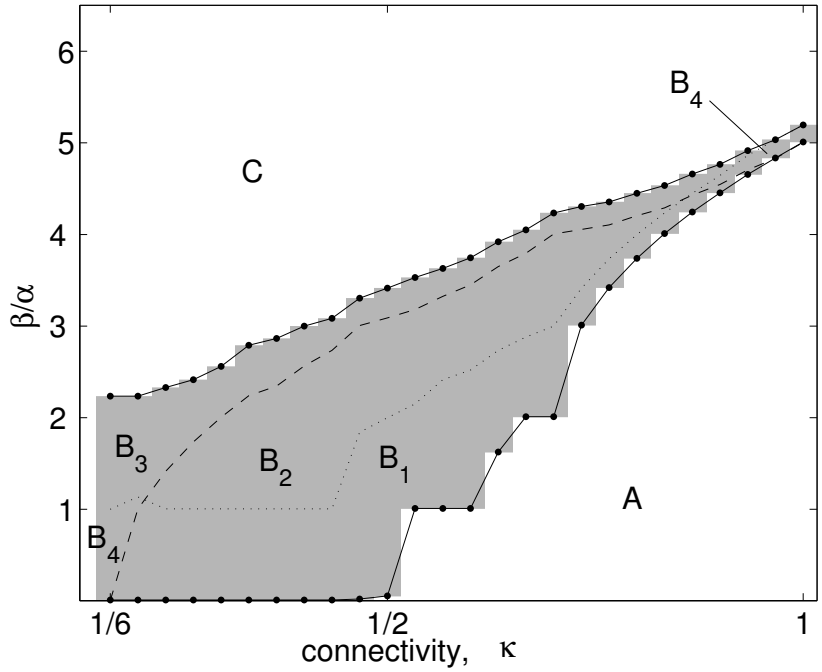


Figure 2.8: Detailed analysis of effect of network connectivity κ and phosphatase activity β/α on stability and signal amplification for networks of size $n = 6$. $\mathbf{B_1}$) Subregion below the dotted and concomitant below the dashed line: Some networks might be unstable and others stable but *all* stable networks are also amplifying. $\mathbf{B_2}$) Subregion above dotted and below dashed line: All sorts of networks can be found here. Stable as well as unstable ones, and stable networks can either be amplifying or dampening. $\mathbf{B_3}$) Subregion above the dotted and concomitant above the dashed line: *All* networks are stable, but only some are amplifying, whereas others are dampening. $\mathbf{B_4}$) Subregion below the dotted and concomitant above the dashed line: All networks are stable *and* amplifying.

for networks of such size. The dashed line denotes the "stability line": Above this line *all* networks are stable. Some of them are dampening whereas others are amplifying. The dotted line displays the "amplification line": Below this line all networks which are

stable are definitely amplifying. The legend of Figure 2.8 describes in more detail by what characteristics the subregions are categorised. There is only a small area within the $(\kappa, \beta/\alpha)$-plane, indicated with $\mathbf{B_4}$, where *all* networks are stable and also amplifying.

As a second structural design property, besides network-connectivity, the number of feedback cycles is considered. An increasing number of cycles have the same tendency as increasing connectivity, namely to make networks less stable. To quantify this tendency and to investigate the influence of the network size n, a scaled measure σ for the number of cycles in a certain network with n nodes is introduced in definition (2.6). In Figure 2.9 the $(\sigma, \beta/\alpha)$-plane is depicted. The abscissa displays the density of cycles, i.e. the

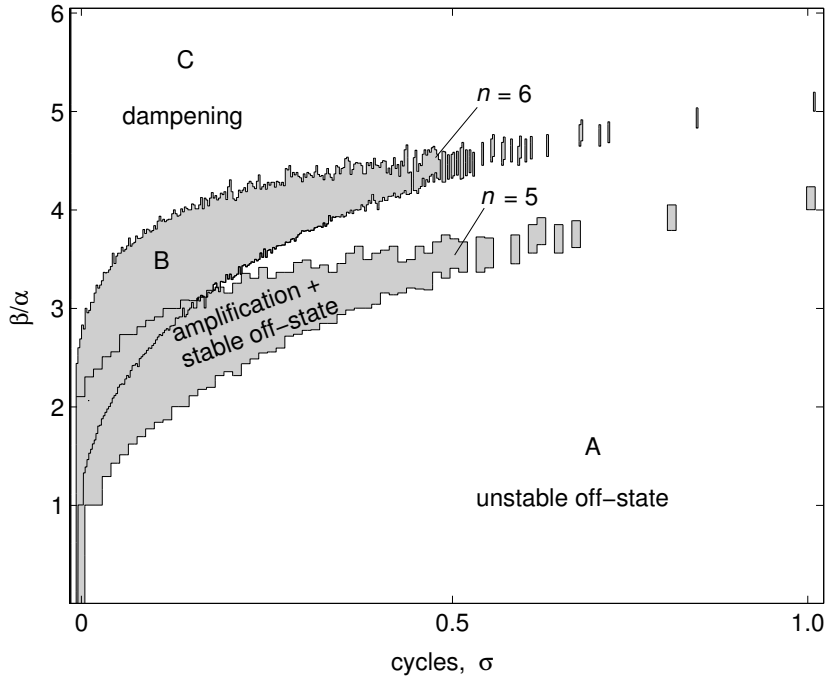


Figure 2.9: Effect of the total number of cycles within a network and phosphatase activity β/α on stability and signal amplification for networks of size $n = 5, 6$. **A**) All networks possess an unstable off-state. **B**) Networks exist which have a stable off-state and exhibit signal amplification. **C**) All networks are stable but dampening.

total number of cycles counted within a network normalised by the maximum number of cycles a network of size n can possess ($\sigma = c/c_{\max}$). Again one identifies three regions **A**, **B** and **C**, representing parameter combinations with unstable, amplifying and dampening networks, respectively. The grey region indicates the parameter region for which networks of size $n = 5, 6$ exist which have a stable off-state and are amplifying

at the same time. It can be seen that the more feedback cycles a network with a given β/α-value contains, the more unstable it gets. This effect is even more pronounced compared to the unstablising tendency of higher connectivities κ. A small number of cycles has already a high destabilising effect on the network. There are, indeed, no records about (direct) positive feedback cycles found in protein kinase networks hitherto. Such feedbacks only exist via transcriptional and translational delay mechanisms, which bring about an additional stabilising effect (Matten et al. [1996]).

2.1.5 Mutations of signalling networks

In the previous sections the structural design of signalling networks has been analysed. The present section treats the question how structural *changes* of signal transduction networks affect their dynamic properties. In principle there are two different scenarios where structural changes in signalling networks are conceivable.

On a short time scale: incorporation or elimination of kinases due to alteration in gene expression or protein degradation.

On a longer time scale: changes in the network structure during evolution. Mutations may lead to a loss or gain of function of kinases to activate certain other kinases. It is also possible that certain mutations of a kinase can stop (or additionally initiate) *other* kinases to act as activators thereof.

Possible transitions from a given network to another network which differs from the initial one in exactly one kinase interaction ($\Delta e = 1$) are of interest. Such structural changes which are called *mutations* in the following define relationships among different networks (mutation neighbours). The presented strategy refers to the concept introduced for metabolic pathways in Ebenhöh and Heinrich [2003]. The basic hypothesis is that those signalling networks (and network populations) which can in principle mutate to a high number of networks with similar dynamic properties, are robust and in that respect they have an evolutionary advantage. For a fixed value of β/α different mutations of a given network can be classified via the dynamic properties of the resulting network. Given the initial network being stable and amplifying, each mutation may lead to a functioning network, i.e. a network that is still stable and amplifying, or to a network that has lost one or both dynamic properties. As an example, Figure 2.10 shows all possible mutations of a particular network (network 1) with four kinases and seven interactions. Out of seven possible mutations of network 1, four take place by losing one interaction (networks 2, 3, 4, 5) and three by gaining one additional activation (networks

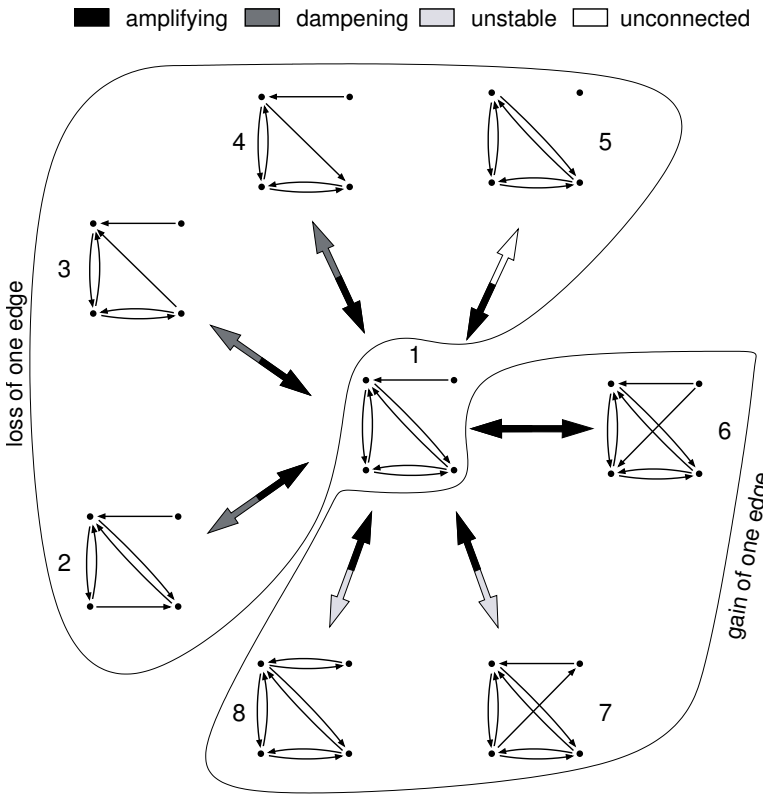


Figure 2.10: Complete set of possible mutations of an initial network with $n = 4$ kinases and their effects on dynamic properties ($\beta/\alpha = 2.1$). The "mutation arrows" are partitioned in different colours according to the properties of the mutation neighbours. Black mutation arrows indicate amplification (network 1 and 6), dark grey dampening (2, 3 and 4), light grey instability (7 and 8) and blank arrows denote a mutation leading to an unconnected network (5).

6, 7, 8). For the given parameters value ($\beta/\alpha = 2.1$), the initial network 1 has a stable off-state and is moreover amplifying, which is indicated by the black coloured part of the mutation arrow. After gaining one additional interaction only network 6 remains amplifying, whereas networks 7 and 8 loose this capacity and become unstable. Instability is indicated by a light grey part of the mutation arrow. This harmful effect is a consequence of the emergence of feedback cycles the impact of which has been analysed in the previous section. Losing an interaction by mutation can also result in dampening networks 2, 3, 4 (dark grey) and in a disconnected network 5 (blank). By identifying possible mutations of all networks of a given size n one arrives at an undirected graph, where the nodes are the signalling networks and the edges are the possible mutations

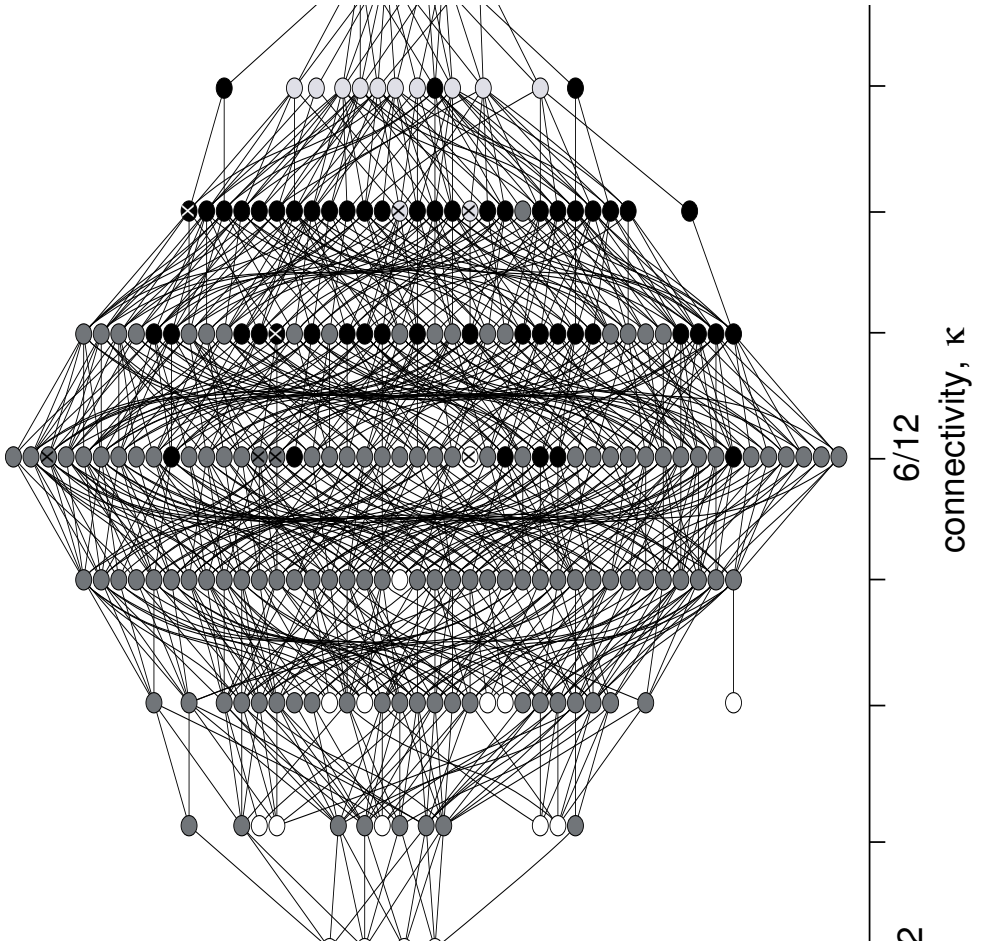


Figure 2.11: Mutation graph for networks of size $n = 4$. The 215 nodes represent networks with four kinases and the undirected edges represent possible mutations. Black nodes indicate stable, dark grey dampening, light grey unstable networks for $\beta/\alpha = 2.1$. Blank nodes denote disconnected networks. Crosses mark the eight example networks from Figure 2.10.

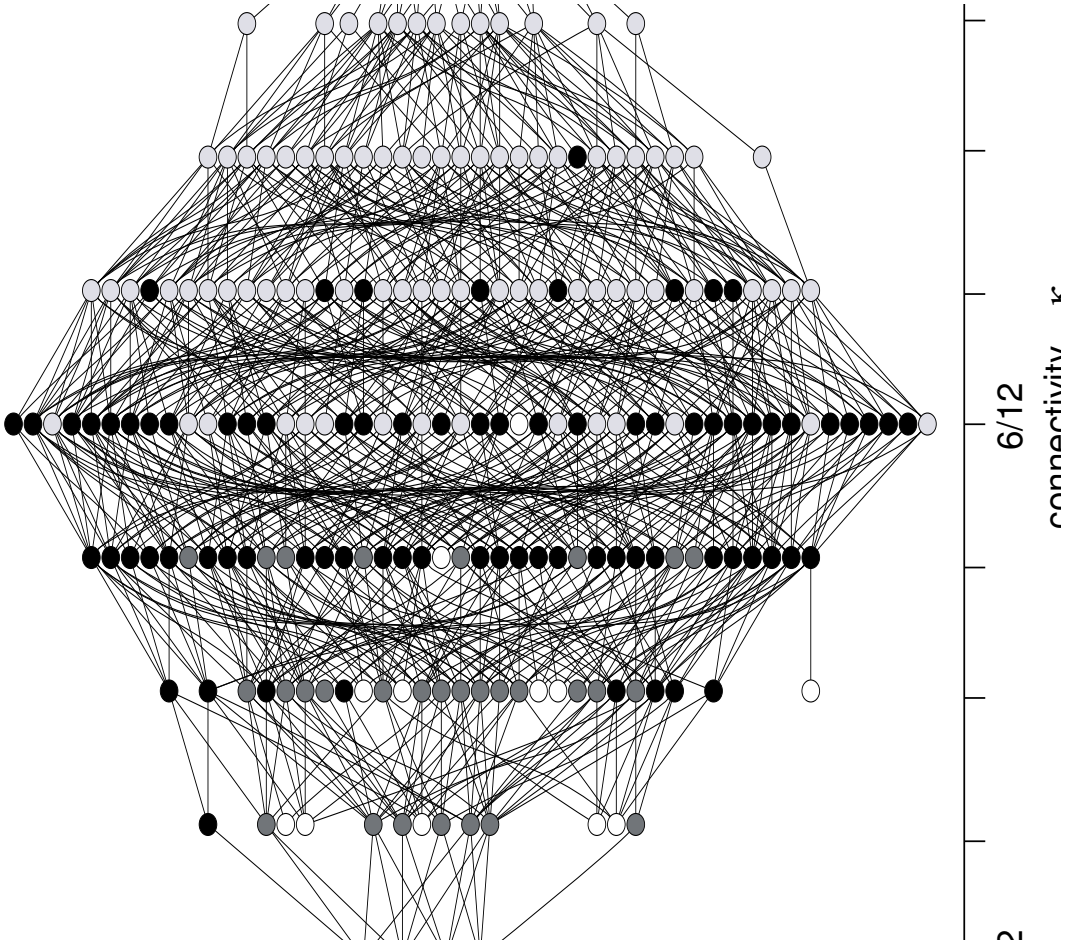


Figure 2.12: Mutation graph for networks of size $n = 4$. The 215 nodes represent networks with four kinases and the undirected edges represent possible mutations. Black nodes indicate stable, dark grey dampening, light grey unstable networks for $\beta/\alpha = 1.5$. Blank nodes denote disconnected networks (c.f. Figure 2.11).

transforming one network into another.

As an example, the mutation relationships of all networks of size $n = 4$ are depicted in Figure 2.11 for $\beta/\alpha = 2.1$. Besides the 199 connected networks also 16 disconnected networks are depicted since in an evolutionary scenario these networks can also be generated if a connected network with four kinases mutates. Therefore 215 networks are considered in the mutation graph which can mutate at least to one other network which is (weakly) connected. The networks are grouped within the mutation graph according to their connectivity κ (abscissa). The smallest connectivities $\kappa = 2/12$ arise after a mutation from a connected network with 3 edges ($\kappa = 3/12$). Like in Figure 2.10 disconnected networks are indicated with blank circles. Black circles indicate signal amplifying networks, dark grey circles signal dampening and light grey unstable networks for the specific phosphatase activity. The eight example networks of Figure 2.10 which may arise after a mutation of network 1 are marked with a cross. As a second example Figure 2.12 displays the situation when the phosphatase activity is lower ($\beta/\alpha = 1.5$) which brings about more networks to be unstable.

Of particular interest is the total number of network designs (connected and unconnected) which can emerge when a *connected* network mutates. The distribution of these numbers for all 199 connected networks of size $n = 4$ is depicted in the lower panel of Figure 2.13. The histogram shows the frequencies of networks with a certain amount of mutation neighbours (abscissa). The lowest possible number of mutation neighbours is 1. One single network shows this mutation character: the fully connected network which can only mutate to the network with $\kappa = 11/12$. A striking feature is that 100 out of 199 networks can mutate to 12 different network designs. This is the maximum number of mutations which can happen in a network with four kinases. Accordingly the maximum number of mutation neighbours that a network of size $n = 4$ can possess is $N_{\max} = n \cdot (n-1) = 4 \cdot 3 = 12$. Networks with small numbers of mutation neighbours are those which show high symmetry concerning the arrangement of their kinase activations. This is because some mutations of such networks result in the same network design.

Mutation graphs consisting of networks of size $n > 4$ show similar properties with respect to "mutation connectivities". Figure 2.13 shows distributions of mutation neighbours for mutation graphs consisting of networks with four, five and six kinases. It should be pointed out that the ordinates which give the frequencies of networks with a certain amount of mutation neighbours have logarithmic scales. It can be seen that irrespective of the size n, most of the networks possess a high variability concerning their mutation potential. Most networks can mutate to all possible mutation neighbours a network of size n can possess. For example $7,536$ networks (out of $9,364$) of

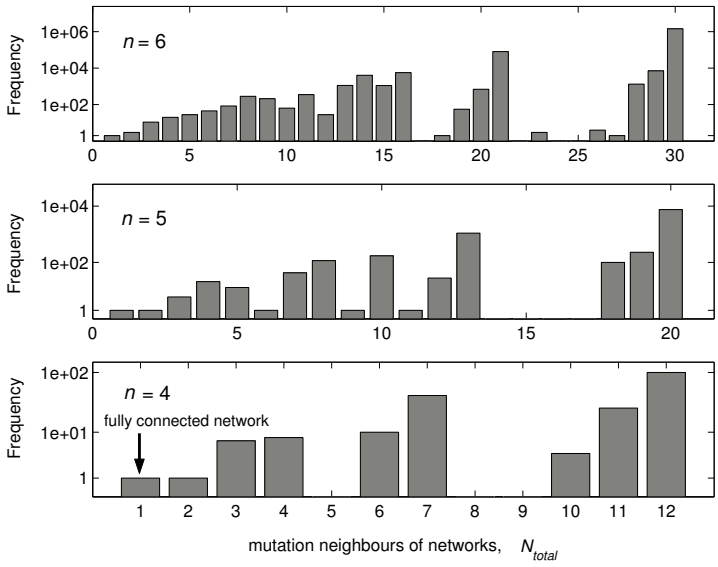


Figure 2.13: Distributions of numbers of mutation neighbours for mutation graphs consisting of $n = 4, 5, 6$ kinase networks. The ordinates have logarithmic scales. On the abscissa fully connected networks mark the lower bound of mutation neighbours. Networks exhibiting the maximum variability, concerning mutations to other networks, mark the upper bound ($N_{\max} = n \cdot (n-1)$).

size $n = 5$ can mutate to 20 different network designs which is more than 80%. In the case of $n = 6$ there are $1,428,852$ networks (out of $1,530,843$) which can mutate to the maximum number of 30 mutation neighbours (93%).

As described in the previous sections quantitatively, dynamic properties of signalling networks change with higher or lower phosphatase activities. For example in Figure 2.11 a specific parameter combination is chosen as $\beta/\alpha = 2.1$. For this particular value a certain set of networks is stable and amplifying (black circles). Highest connectivities within this set of networks are found to be $\kappa = 0.5$. Increasing β/α will make networks with higher connectivities κ become stable and instantaneously amplifying, whereas networks with lower κ will become dampening. Together this shifting of the value for β/α results in a maximum for the number of amplifying networks at a certain phosphatase activity. Figure 2.14A shows how the number of amplifying (n_{A}) and the number of dampening (n_{D}) networks depend on β/α. The dashed line marks the total number of 199 networks with $n = 4$. For phosphatase activities up to $\beta/\alpha = 1.0$ there is a constant amount of stable and at the same time amplifying networks ($n_{\mathrm{A}} = 24$) which are exclusively those without any feedback cycles (c.f. Figure 2.7). As the plotted

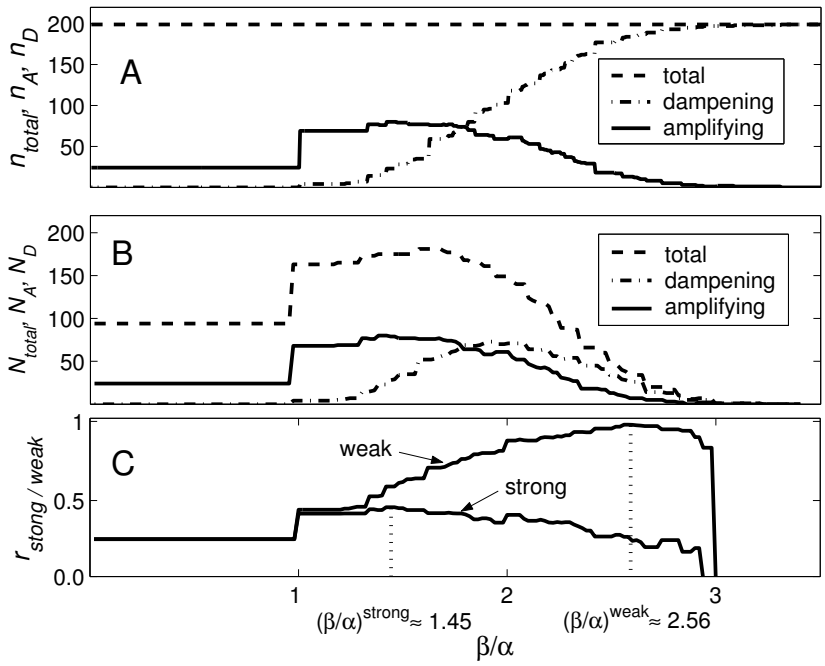


Figure 2.14: Numbers of amplifying and dampening networks, mutation neighbours of exclusively amplifying networks and structural robustness of 4 kinase signalling networks. These numbers depend on the phosphatase activity β/α. A) Total number of networks (= 199, dashed line). Number of amplifying (n_{A}, straight line) and dampening (n_{D}, dash-dotted) networks. B) Mutation neighbours of the ensemble of amplifying networks. Total number (N_{total}, dashed line), number of amplifying (N_{A}, straight line) and dampening neighbours (N_{D}, dash-dotted). C) Strong global robustness r_{strong}: fraction of mutation neighbours which are stable and amplify signals. Weak global robustness r_{weak} : fraction of mutation neighbours which possess at least stable signal-off states.

straight line shows, most networks ($n_{\mathrm{A}} = 80$) are stable and amplifying for $\beta/\alpha \approx 1.5$. For $\beta/\alpha > 3.07$ all networks are dampening.

A concept for structural robustness of a certain network is introduced by considering the dynamic properties of all mutation neighbours of this network. Appropriately the considered network is assumed to function properly, i.e. it is stable and amplifying, before the mutation occurs. In that respect a network is said to be structurally robust if a high fraction of mutation neighbours is also stable and amplifying. The robustness of a network depends on the phosphatase activity β/α. Instead of considering the robustness of single networks, a *global robustness* is introduced. For this purpose all mutation neighbours of *all* stable and amplifying networks (n_{A}) at a certain β/α-value

are accounted for. N_{A} (N_{D}) denotes all stable and amplifying (dampening) neighbours which can be generated by a mutation of any of the n_{A} networks. Figure 2.14B shows how N_{A} (straight line), N_{D} (dash-dotted) and the total number of neighbours N_{total} (dashed) of the n_{A} networks depend on β/α. These numbers have maximum values for intermediate phosphatase activities, since n_{A} is also highest for these β/α-values. In the range $0 < \beta/\alpha < 1$ a total number of $N_{\mathrm{total}} = 94$ different network designs can be generated from mutations of the $n_{\mathrm{A}} = 24$ amplifying networks. Out of these 94 mutation neighbours 70 networks are unstable, whereas $N_{\mathrm{A}} = 24$ remain stable and, moreover, amplifying. Thus, in this range of phosphatase activities all stable and amplifying networks possess at least one mutation neighbour which is also stable and amplifying. Therefore there is no "isolated" stable and amplifying network in the mutation graph, i.e. every such network can be "reached" by one mutation $n_{\mathrm{A}} = N_{\mathrm{A}} = 24$. However, β/α-values exist for which not all stable and amplifying networks can mutate to neighbours which remain amplifying, i.e. such networks are isolated from other amplifying networks within the mutation graph for these particular β/α-values.

If N_{A} and N_{D} are related to the total number of mutation neighbours N_{total}, a measure for (global) strong and (global) weak robustness of the set of n_A networks is obtained as follows:

$$\begin{aligned} \text{strong robustness:} \quad r_{\mathrm{strong}} &= \frac{N_{\mathrm{A}}}{N_{\mathrm{total}}}, \\ \text{weak robustness:} \quad r_{\mathrm{weak}} &= \frac{N_{\mathrm{D}} + N_{\mathrm{A}}}{N_{\mathrm{total}}}. \end{aligned} \tag{2.17}$$

In both cases the robustness is defined for a certain β/α-value. These definitions reflect how likely it is on average for a set of proper functioning, i.e. stable and amplifying networks, to remain functioning after a mutation. In Figure 2.14C it is shown how weak and strong robustness depend on β/α. The value for strong robustness of networks with 4 kinases in the range of $0 < \beta/\alpha < 1$ is $r_{\mathrm{strong}} = 24/94 = 0.2553$. Because dampening networks are lacking in this range it holds $r_{\mathrm{strong}} = r_{\mathrm{weak}}$. In the case of strong robustness one finds the most robust scenario for $\beta'_{\mathrm{max}} \approx 1.45$. This is similar to the β/α-value for which a maximum number of stable and amplifying networks n_A can be found at about $\beta/\alpha = 1.5$ (see maximum of straight line in panel A). However in the case of weak robustness which refers to all stable mutation neighbours, which might be either amplifying *or* dampening, the most robust scenario is found at $\beta'_{\mathrm{max}} \approx 2.45$.

This value is significantly higher than the one, for which one finds a maximum number of amplifying networks.

2.2 Design and dynamics of the kinase network from a signalling database

In the previous sections, general relations between the structure and dynamics of a large class of signalling networks consisting of a relatively low number of kinases has been considered. Some estimations could be made concerning the expected connectivity and the number of feedback cycles of larger kinase networks. Now, signalling networks as can be found in living cells are examined with regard to these estimations.

The network depicted in Figure 2.15 has been deduced from the database TRANSPATH. It includes 86 kinases. The actual number of entries for the keyword "kinase"is much higher (1100). However, many of these entries concern complexes of kinases with other proteins or are ortholog kinases in different species. A comparison of different kinomes, that is, the totality of protein kinases in cells, shows that specific kinase families, and sometimes individual kinases, are conserved even between very divergent organisms, a property which highlights the basic roles of kinases in many cellular processes. An even higher overlap of kinases is assumed for different mammalian kinomes (Manning et al. [2002]).

Additionally to the kinases one obtains from the database the information of mutual activations between kinases. They are annotated as "Reaction downstream" and "Reaction upstream". A more detailed analysis reveals that some of the interactions remain unclear, in particular concerning the question of direct or indirect interaction mode. Taking into account the quoted literature and more recent publications to minimise uncertainty, one arrives at 171 direct mutual activations together forming a (weakly) connected kinase network. Isolated kinases which do not activate other kinases or being itself activated are not taken into account.

The database TRANSPATH also contains information on small G-proteins which may switch between an inactive GDP- and an active GTP-form. Adding these G-proteins and their activating effects to the kinase network results in the network shown in Figure 2.16. Compared to the network containing only kinases the number of compounds is enlarged to 94 and the number of interactions to 199.

The networks have been visualised by using the program graphviz[1], comprising a

[1] http://www.research.att.com/sw/tools/graphviz/

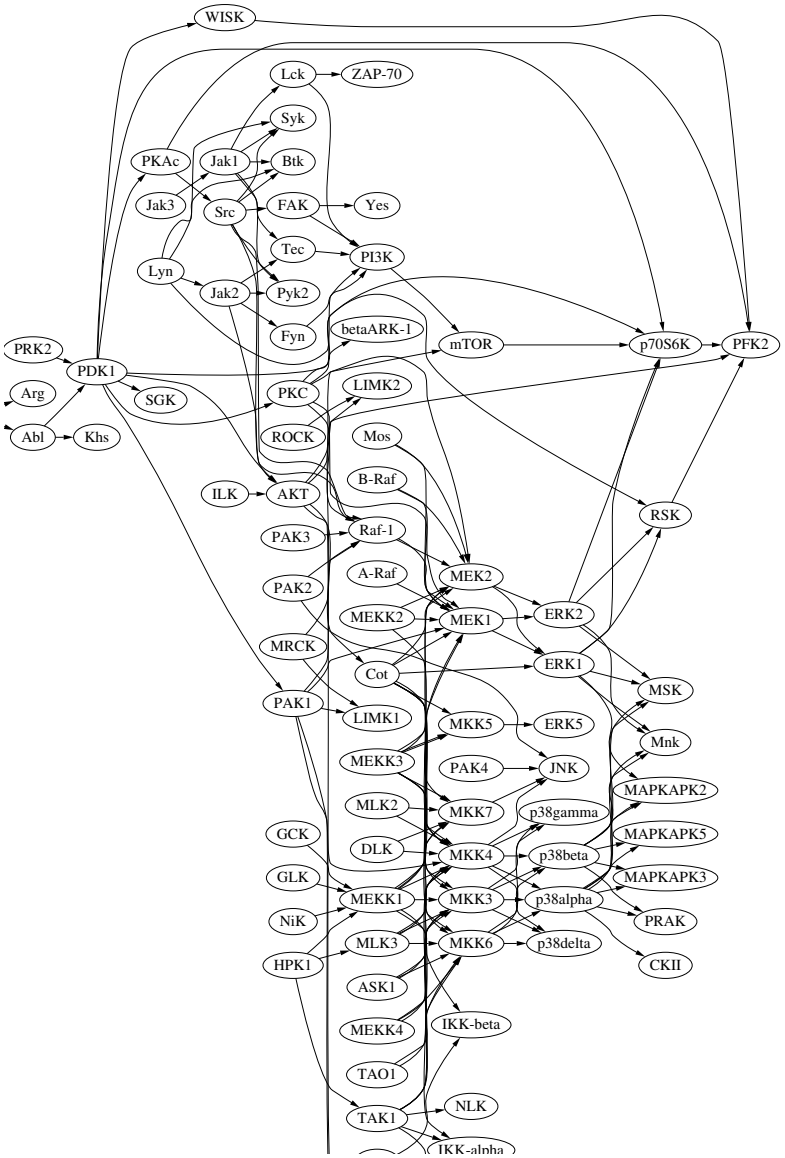


Figure 2.15: Signalling network from the database TRANSPATH. It contains 86 kinases (encircled nodes) which, according to present knowledge, perform 171 mutual activations (edges).

set of graph drawing tools. The layout algorithm produces layered drawings of directed graphs. As far as possible edges are drawn in the same direction (e.g. top to bottom, or left to right) and edge lengths are reduced. Moreover, crossing of edges is avoided when possible. Visualization by this program reveals already several interesting characteristics

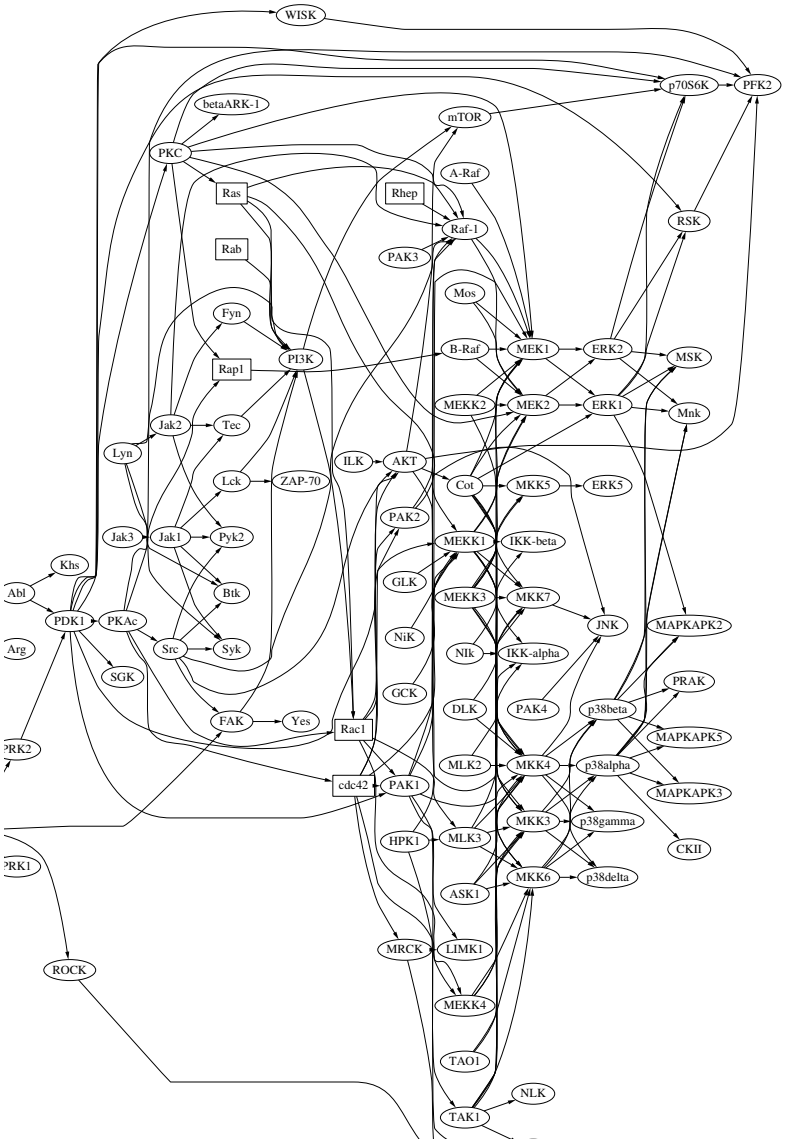


Figure 2.16: Signalling network deduced from the database TRANSPATH. It contains 86 kinases (encircled nodes) and 8 small G-proteins (in boxes) which, according to present knowledge, perform 199 mutual activations (edges).

of the network such as symmetry, parallelism or regularity. In the kinase network depicted in Figure 2.15 one may perceive a layered structure reflecting groups of upstream and downstream kinases. For example, well known components like MAP kinases (e.g. ERK, JNK, p38), MAPK kinases (MEK, MKK) and MAPKK kinases (Raf, MEKK)

are assigned to three consecutive layers.

2.2.1 Feedback cycles, connectivity and pathway lengths

Closer inspection of the kinase network reveals that it does not contain cycles of any length, which is, in view of the high number of interconnections, a rather remarkable property. Inspection of the network containing small G-proteins shows that even this network still contains no cycles.

To asses if this is a special property of the signalling network under consideration, the probability that corresponding random networks do not contain cycles is evaluated. To allow for an appropriate comparison random networks which have the same single-node characteristics as the kinase network in Figure 2.15 are considered (null hypothesis). Specifically, each node of the randomised networks has the same number of incoming and outgoing edges as the corresponding node in the database network (see Figure 2.17). For generating these random networks the method described in Milo et al. [2002] is used. The idea behind this procedure is to use the real network as a starting point and to exchange for two randomly chosen edges the target nodes or the nodes of their origin. Applying this reshuffling process many times, the network structure changes considerably but the numbers of incoming edges and outgoing edges are for each node the same as for the original network.

In the present study $100,000$ different random networks are generated from a copy of the TRANSPATH network. In each case the edges are reshuffled $5,000$ times. This high number of exchanges ensures that the networks undergo sufficiently large changes and that the resulting networks are, despite of the conservation of the node degrees, not biased towards the structure of the initial network. Whereas it is rather cumbersome to determine the exact number of cycles in each case, one may easily calculate the number of networks containing *no* cycles at all. One obtains the result that within the full ensemble of $100,000$ networks under consideration only 37 networks do not contain cycles. This suggests that the probability of finding a **D**irected **A**cyclic **G**raph (DAG) among the randomised networks is extremely low. A rough estimate might yield $P_{\mathrm{DAG}} \approx 0.04\%$ for this probability. One may conclude therefore, that the absence of cycles is a very striking feature of the transpath network

For further structural characterisation the connectivity κ of the two kinase networks from the database are calculated. According to formula (2.5) one obtains for the network with only kinases and the one with additionally G-proteins $\kappa = 0.0234$ and $\kappa = 0.0228$, respectively. This means that each single component has on average about two inter-

action partners. These low numbers support our hypothesis that realistic signalling networks of large size must be characterised by a low connectivity in order to avoid autoactivation in the absence of any receptor stimulus.

A more detailed characteristic of the interconnections concerns the number of activations received or performed by any given kinase. For example PDK1 is activated by two upstream kinases and itself activates eight other kinases (see Figure 2.15). MEK1 behaves in an opposite way: It has ten upstream and only two downstream kinases. The diagram in Figure 2.17 shows how many kinases are found having given numbers of upstream and downstream interactions. There is a recognizable tendency of the kinases

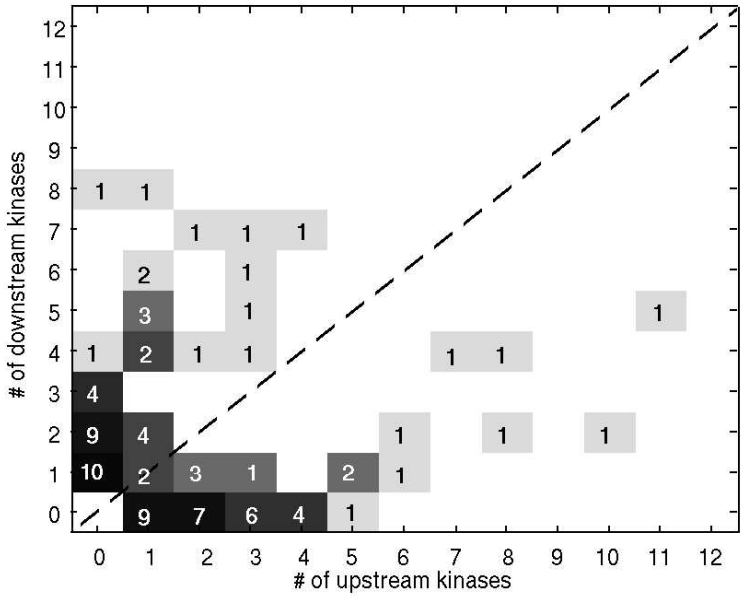


Figure 2.17: Unspecific interactions within the kinase network of Figure 2.15. The entries give the numbers of kinases characterised by given numbers of input and output activations.

to divide into two groups. Some act predominantly as signal distributors having many downstream kinases as substrates but few activating kinases. They are represented by numbers above the diagonal in the diagram of Figure 2.17). Kinases of the second group act as signal receivers having more upstream kinases than downstream targets (numbers below the diagonal).

A remarkable property of MAP kinase pathways is that they consist of three to four consecutive levels of a phosphorylation cascade. It is an intriguing question, if such a

pathway length is typical also for other pathways which can be identified in the whole network. As pathways all routes of activations are considered starting from an input kinase which do not have activating upstream kinases and leading to an output kinase which does not activate other downstream kinases. The network contains 25 input kinases and 27 output kinases constituting 675 pairs of kinases between which routes could in principle exist. Inspection shows that not all output kinases are reachable by routes from any input kinase and that routes (of various lengths) only exist between $N_{\mathrm{io}} = 222$ input/output kinase pairs. The total number of routes amounts to $P^{\mathrm{total}} = 1272$ meaning that there are, on average, six routes between each pair of connected kinases. As a specific feature the lengths of the shortest pathways between any two connected kinases are calculated. The histogram in Figure 2.18A shows how many shortest routes of a given length L exist. It can be seen that no route is longer than

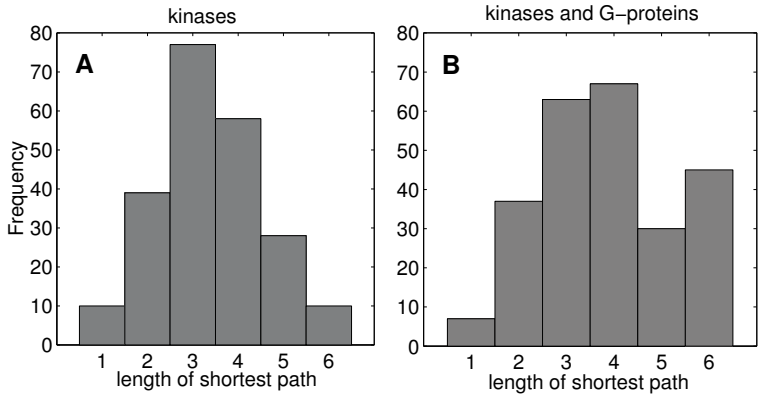


Figure 2.18: Lengths distribution of shortest pathways between input and output kinases of (A) the kinase network in Figure 2.15 and (B) the network of kinases and small G-proteins in Figure 2.16.

six and there are a considerable number of routes consisting of only one step. Routes of length $L = 3$ occur more often than shorter or longer routes. The typical lengths of routes in this network correspond therefore to that of MAPK-pathways. In Figure 2.18B the distribution of pathway lengths is shown for the network depicted in Figure 2.16 which also involves G-proteins. Although the number of additional components is only slightly increased by the inclusion of G-proteins (from 86 to 94), there is a rather strong effect on the lengths' distribution. Compared to the kinase network the number of the longest routes with six activations is increased by a factor of more than four. Moreover, the most frequent route length is shifted from three to four. This corresponds, for

example, to the situation within the MAPK pathways whose first kinases are activated by a G-protein.

Among the 222 connected pairs of kinases in the TRANSPATH network of Figure 2.15 there are 57 pairs which are connected by only one route. In another class which contains 134 pairs there exist several routes but all of them are of equal length, that is, all of these routes represent shortest routes. There exist only 31 pairs which are connected by routes of different lengths. Within the latter class the longest routes between two kinases comprise 10 activations steps. These results indicate that kinases are typically interconnected by shortest routes. To compare this structural feature with the distribution of pathway lengths of other networks for each connected pair i of kinases the fraction of shortest routes is introduced:

$$f_i = \frac{s_i}{P_i}, \tag{2.18}$$

where s_i and p_i denote the number of shortest routes and the total number of routes, respectively, between the two kinases of pair i. The fraction of shortest routes among all routes within the network is defined as

$$F = \frac{1}{N_{\mathrm{io}}} \sum_{i=1}^{N_{\mathrm{io}}} f_i, \tag{2.19}$$

with $0 < F \leq 1$. High values of F indicate that a network is rather homogenous in the sense that it consists mainly of shortest routes between connected kinases. For the TRANSPATH kinase network one obtains $F \approx 0.86$.

Pathway lengths can also be calculated for random networks. For the corresponding study 100 networks randomised in the way explained above are considered. Since the original network does not contain cycles randomisation is performed under the constraint that the resulting networks also do not contain cycles. Since the degrees of the nodes are preserved all these networks have the same number of input kinases (25) and output kinases (27) as the TRANSPATH kinase network. However, the number of connected pairs is generally different as depicted in the histogram shown in Figure 2.19A. The range of the N_{io}-values lies between 168 and 349, and on average the networks contain approximately 253 pairs of connected kinases. For a comparison with the distribution of N_{io}-values the histogram in Figure 2.19B depicts the distribution of the total numbers of routes which can be found in the random networks. These P^{total}-values range from 680 to 5659. The fraction of shortest routes for each randomised network is also calculated leading to the result depicted in Figure 2.19C. Compared with the TRANSPATH network (whose F-value is indicated by an arrow) the fraction of shortest routes within

randomised networks is generally much lower, meaning that random networks show a much more diverse distribution of pathway lengths.

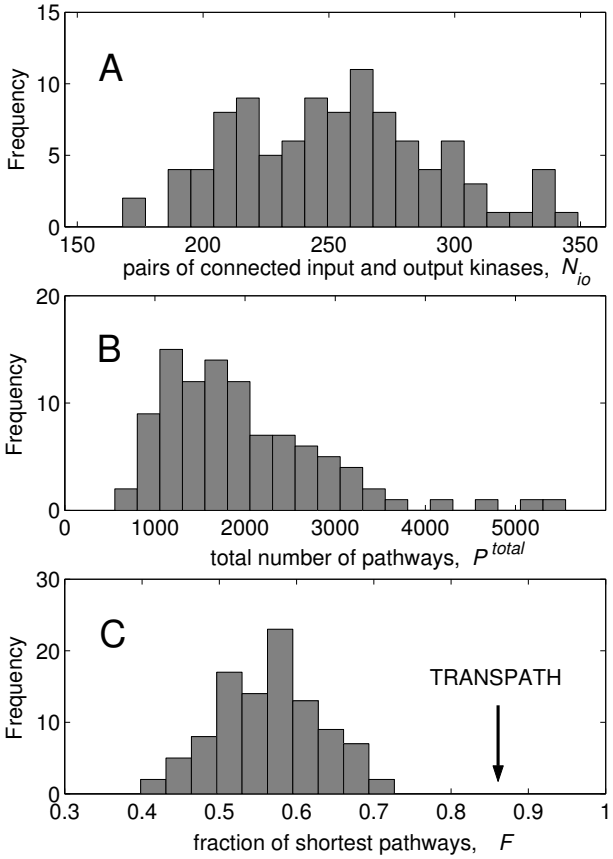


Figure 2.19: Pathway composition in 100 random networks. A) Distribution of numbers of input/output kinase pairs N_{io} which are connected by routes. N_{io} is equal to the number of shortest pathways within a network. B) Distribution of total numbers of pathways P^{total} found in each random network. C) Distribution of fraction of shortest pathways F.

As special examples three randomly generated networks, with the lowest, a medium and the highest N_{io}-value (168, 253, and 349, respectively) are considered in more detail. The histograms in Figure 2.20A–C show the distributions of shortest routes of these particular random networks. Interestingly, the typical lengths of shortest routes is again $L = 3-4$ and also the shape of the distributions are similar to that of the TRANSPATH network. However, shortest pathways which are longer than $L = 6$ also exist. The longest routes which one finds therein comprise 10, 14, and once more 14 activations

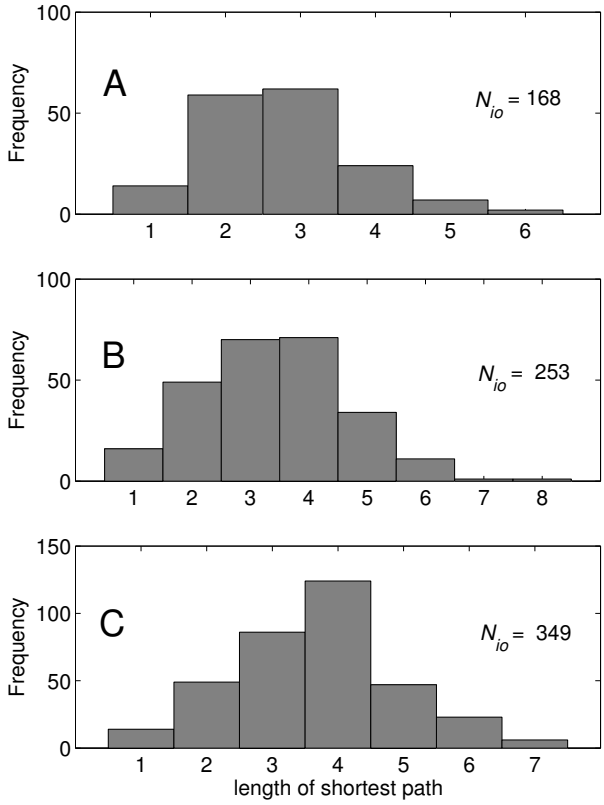


Figure 2.20: Lengths distribution of shortest pathways between input and output kinases of three example random networks with 86 kinases and 171 activations. The total number of shortest pathways amounts to 168 (A), 253 (B), and 349 (C), respectively.

steps, respectively. The fact that the random networks involve also longer routes is reflected by the corresponding fractions of shortest routes, $F = 0.66$, $F = 0.53$, and $F = 0.44$, respectively, that is of values which are much smaller than that for the TRANSPATH network. The results suggest that real kinase networks have been selected for a more uniform distribution of pathway lengths favouring the shortest routes.

In the next section 2.2.2 the 1272 pathways which are found within the TRANSPATH kinase network are analysed in more detail with respect to crosstalk.

2.2.2 Signalling crosstalk

It is a known phenomenon that although different MAP kinase cascades run in parallel, there is a considerable degree of crosstalk between them, which creates multiple opportunities for modulating or fine-tuning responses to different signals (Cowan and

Storey [2003]). Moreover, specificity of the MAP kinase signalling pathways is greatest at the level of specific MKK activation of individual MAPKs, where there is the least amount of crosstalk. There is often considerable crosstalk at the MAPK substrate level, e.g. transcription factors are phosphorylated by many MAP kinase signalling pathways. The MAPKKKs are also involved in crosstalk, although much remains to be elucidated at this level. Figure 2.21 outlines various known crosstalk interactions among MAP kinase pathways. The differences, regarding the amount of crosstalk within consecutive

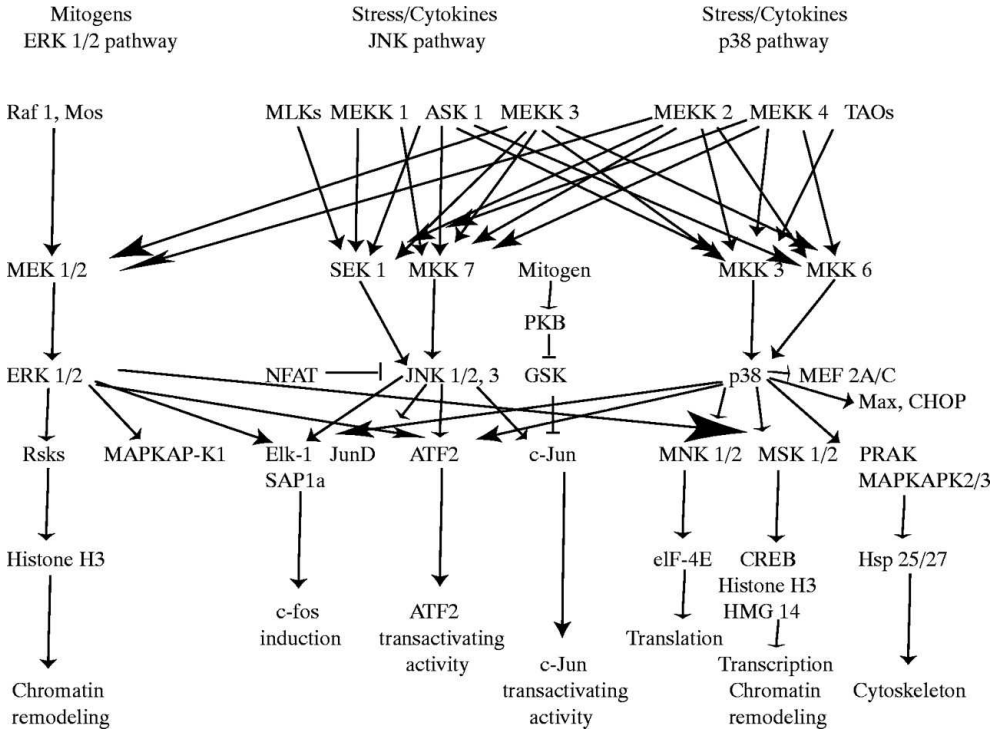


Figure 2.21: Summary of MAPK signalling pathways, showing mutual crosstalk (figure from Cowan and Storey [2003]). A striking feature is the different magnitude of crosstalk within the three levels of the cascades.

levels, can be clearly observed. Instead of focusing on this particular property, a more general question is discussed in the present section by introducing a measure to quantify crosstalk between two distinct pathways. The aim is to investigate, if the global distribution of crosstalk strengths, which is encountered in the TRANSPATH network, differs significantly from the distributions in random networks. Therefore, all 1272 pathways are considered pairwise. However, some of these pathways contain common kinases. Crosstalk can be defined most conveniently between pathways which do not contain a common kinase, i.e. which are fully disjoint. Provided that each considered pair does

not have a kinase in common, one ends up with $670,282$ pairs to account for. This number is about 20% lower than the total number of pathway pairs (including also pairs of overlapping pathways) which amount to $1/2 \times 1272 \times 1271 = 808,356$.

The strength of crosstalk between a pair of pathways (P1, P2) is measured by counting all cross interactions (edges) between the kinases of P1 and P2. This method is illustrated with two disjoint example pathways P1 and P2 in Figure 2.22A. The pathways have lengths of 3 and 2, respectively. The dashed arrows indicate two existing cross-activations between P1 and P2. There is a maximum number of activations which are in principle possible between two distinct pathways of certain size. This can be seen in Figure 2.22A, where the light grey arrows mark additional 6 cross-activations which would also be possible between P1 and P2. These eight in principle possible cross-activations constitute the maximum number of activations between disjoint pathways of length 2 and 3.

To obtain a scaled measure for crosstalk, the ratio of existing interactions to the maximum interactions which are in principle possible between two pathways is considered:

$$\eta = \frac{\text{actual number of interactions between two distinct pathways}}{\text{total number of possible interactions between two distinct pathways}}. \qquad (2.20)$$

Using this definition for the crosstalk between the example pathways in Figure 2.22A one calculates $\eta_{\mathrm{P1,P2}} = 2/8 = 0.25$. The distribution of crosstalk η of the TRANSPATH kinase network is shown in Figure 2.22B. For comparison η-values of random networks without cycles are calculated. As an example, the distribution of η values of the single random network with $N_{\mathrm{io}} = 253$ – containing a total number of 1632 pathways – is depicted in Figure 2.22C. Both distributions mainly have η-values between 0 and 0.5. Although the number of input and output kinases is kept constant, random networks tend to have many more pathways (higher P^{total}-values) and accordingly more pairs of pathways have to evaluate for crosstalk. η-distributions are calculated for 100 random networks. However, since they look similar in their shape (cf. Figure 2.22C), the η-distribution of only one random network is shown as an example distribution. Two main characteristics might be recognised in the crosstalk distribution of the kinase network from TRANSPATH:

1) One finds a comparably high number of pathway pairs which do not show any crosstalk at all ($\eta = 0$).

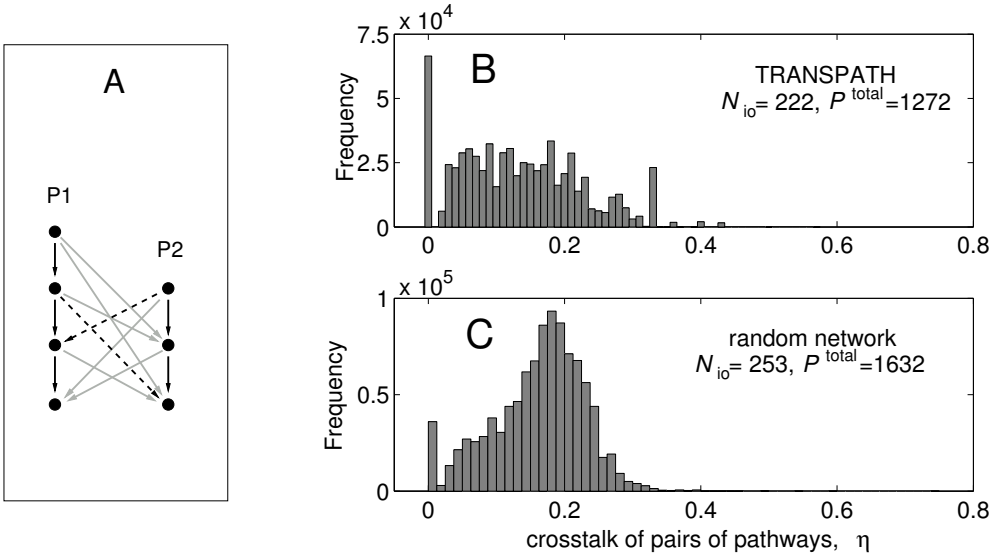


Figure 2.22: Crosstalk between two pathways. A) Example: Disjoint pathways P1 and P2 (black solid arrows) exhibit two crosstalk interactions (dashed arrows). The maximum number of crosstalk activations (dashed *and* grey arrows) which can in principle be encountered between pathways of these lengths is 8 ($\Rightarrow \eta_{\mathrm{P1,P2}} = 2/8 = 0.25$). B) Distribution of crosstalk η within the kinase network of Figure 2.15: 1272 pathways give rise to $670,282$ disjoint pairs of routes to evaluate. C) Distribution of crosstalk within a single randomly generated network without cycles and $N_{\mathrm{io}} = 253$. It contains $P^{\mathrm{total}} = 1632$ pathways giving rise to $1,010,082$ disjoint pairs of routes and the same number of η-values.

2) There is a rather monotonous decrease of the number of those pairs which exhibit higher crosstalk strengths. In contrast, random networks exhibit considerably more intermediate crosstalk, i.e. around a certain mean value of η.

These calculations show that design principles with respect to pathway crosstalk in the database network are considerably different from the structure of random networks. Cross-interactions in "real" kinase networks are not arranged arbitrarily, but they seem to be rather "designed".

Hitherto, random networks have been used as a "null hypothesis" to assess the significance of structural features of the TRANSPATH kinase network. In the next section 2.2.3, random networks – also based on the same node degrees as the TRANSPATH network – are interpreted differently, namely in association with a possible evolutionary strategy of kinase networks to avoid the emergence of feedback cycles.

2.2.3 A possible strategy to avoid cycles

Besides the basic constraint which preserves the degrees of all nodes in random networks, it is moreover interesting to apply additional constraints with respect to certain *small* substructures of the initial kinase network. For example in Milo et al. [2002], randomised networks are used to calculate the significance of certain subgraphs of size m in a real network by preserving the same number of appearances of all $(m-1)$-node subgraphs found in the real network. This strategy ensures that a high significance is not assigned to a pattern only because it has a highly significant subpattern. It emphasises that a possibly frequent occurrence (or absence) of larger substructures in the reference network is not the result of a combination of smaller substructures.

In section 2.1.4, the dangerous effect of cycles – particularly for large signalling networks – is studied analytically. Therefore, during evolution a kinase network might aim to avoid such direct positive feedback mechanisms. To test if the TRANSPATH kinase network has the intrinsic *potential* to avoid such cycles – and if so, by which mechanisms – random networks based on the TRANSPATH kinase network are generated by applying further constraints. As a mechanism to avoid cycles entirely, the impact of inhibiting the emergence of small cycles in random networks is taken into account. Claiming that small cycles must not occur in the considered random networks will certainly increase the probability of finding networks *without any* cycles (P_{DAG}). For example, claiming additionally that the random network algorithm only generates networks without feedback cycles of length two ($C_2 = 0$) increases P_{DAG} to about 0.6%. Table 2.3 lists how P_{DAG} depends on the constraints on small cycles for random networks generated on the basis of the kinase network from TRANSPATH.

constraints:	only node degrees are conserved	degrees & $C_2 = 0$	degrees & $C_2 = C_3 = 0$	degrees & $C_2 = C_3 = C_4 = 0$
P_{DAG}	$\approx 0.04\%$	0.6%	7%	61%

Table 2.3: Probabilities of finding a structural design without any feedback cycle (P_{DAG}) among randomised signalling networks. A copy of kinase network from TRANSPATH is taken to generate $100,000$ random networks, i.e. the node degrees of all 86 kinases are in every case the same as in Figure 2.17. Besides this constraint, all random networks are generated with four different additional constraints by excluding small cycles: 1) No constraints on cycles, 2) C_2 cycles are excluded, 3) C_2, C_3 cycles are excluded, 4) C_2, C_3, C_4 cycles are excluded.

In contrast to the calculations where random networks are taken as a null hypothesis for an estimation of how likely it is to find cycles or certain pathway lengths in random networks, the interpretation is different in the present case: The random networks with additional constraints to avoid small cycles may be interpreted as carrying the same potential as the TRANSPATH network (on the basis of the degree distribution) to avoid cycles entirely by a specific mechanism.

The results can be summarised as follows: The more consequent smaller cycles are excluded from random networks with the same degree distribution as the network from TRANSPATH, the more likely it becomes that these networks can avoid cycles entirely. The calculations so far show that for signalling networks which aim to avoid cycles, a good strategy is to avoid in particular small feedback cycles of length $2-4$ during an evolutionary process. This would dramatically increase the probability that the evolving kinase network remains without any cycles. Many scaffold proteins are thought to arrange three consecutive kinases properly to allow for a signal to be transduced (Yoshioka [2004]). Such an arrangement might ensure that the last scaffold-bound component is prevented from activating the upstream components, thereby avoiding autoactivation.

Since the design of kinase networks have evolved as a result of an interplay with their dynamic properties, section 2.2.4 investigates the dynamic behaviour of the TRANSPATH network.

2.2.4 Dynamic behaviour of network components

In the present section, the dynamic properties of this network are analysed. Thereby the question of how the signal spreads out through the network is of interest. For this purpose, the same kinetic model as in section (2.1) which is based on ordinary differential equations and uses mass action kinetics (see Eqs. 2.3) is used. As the network contains no cycles, its signal-off state is dynamically stable, irrespective of the kinetic parameter β/α. To assess the dynamic characteristics of the network, one has to choose an appropriate receptor attachment to certain input kinases. For full activation of the network one must attach receptors to all input kinases. The parameter for the receptor decay is chosen as $\lambda = 1$. Parameters describing the ratio between the phosphatase and kinase activity are all set to be equal ($\beta/\alpha = 1.0$) and a situation for weakly activated pathways is assumed ($X_i^* \ll C_i$). Under these conditions, the amplitudes S_i, signalling durations ϑ_i and signalling times τ_i of the activation profiles of all 86 kinases are calculated. All input kinases have the same time profile and therefore the same dynamic key values. Because of the specific design, some other kinases within the network have the same amount and

kind of upstream signalling routes leading to them which brings about identical profiles for these kinase. One ends up with 42 discriminable time profiles and consequently (τ, S)-values and (ϑ, S)-values, respectively. The result for the (τ, S)-values is shown in Figure 2.23. The abscissa describes the signalling time τ and the ordinate the signal

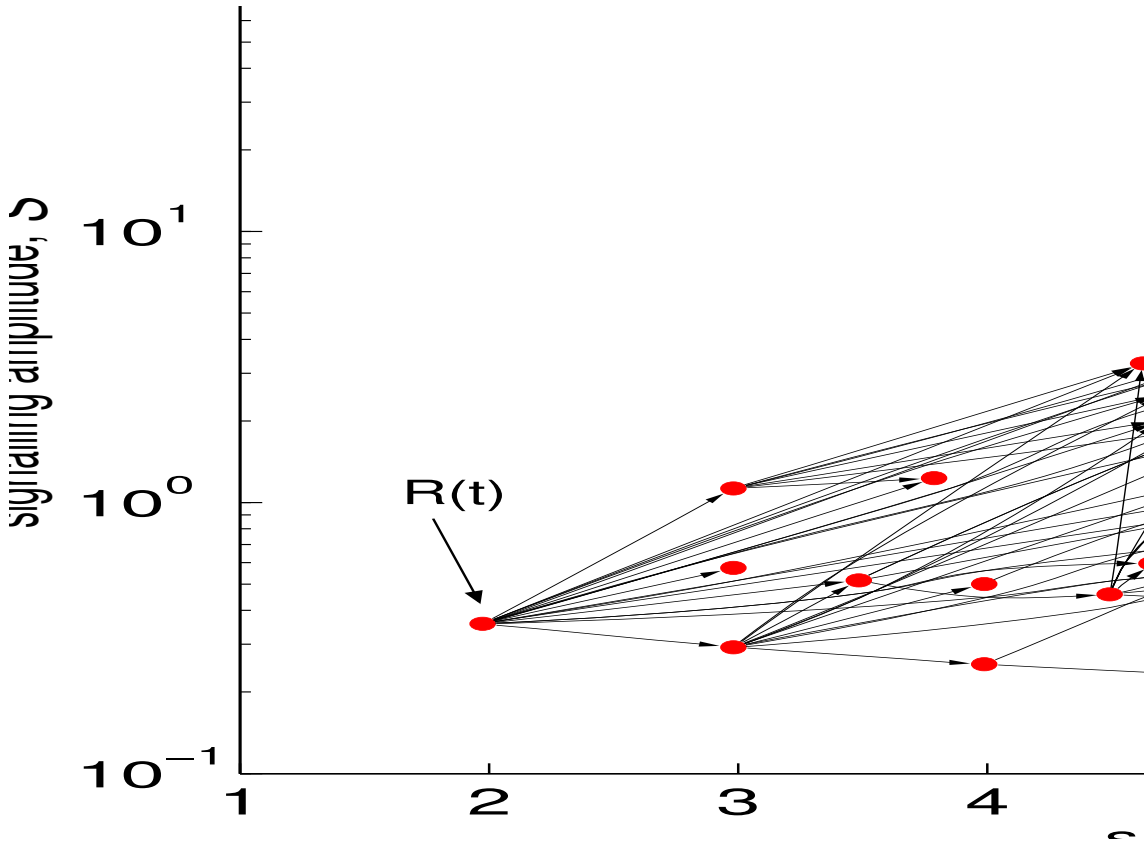


Figure 2.23: Dynamic properties of the kinase network from TRANSPATH. Abscissa: signal time τ. Ordinate: signalling amplitude S. Receptors $R(t)$ are attached to all 25 input kinases. Example kinases with significantly high amplitudes are indicated in the graph: p70S6K, MAPKAPK2, Mnk, MSK, PFK2.

amplitude S of the kinases. In this depiction, there is one single source node, standing for the (τ, S)-values of *all* 25 input kinases where a receptor is attached. Some of the other nodes also stand for multiple kinases which have the same profiles and thereby (S, τ)-values.

Whereas the signalling times are confined within one order of magnitude $2 \lesssim \tau \lesssim 8$, the amplitudes vary between $0.4 \lesssim S \lesssim 40$. Kinases with significantly high amplitudes are marked in the figure. As know from textbooks they are involved in metabolic

processes and the translation machinery. For example, the ribosomal kinase p70S6K is a positive regulator of the S6 ribosome subunit. Mnk phosphorylates and thereby activates Eukaryotic translation initiation factor 4E (eIF4E) which binds to the mRNA and brings the mRNA into a complex with other protein synthesis initiation factors and ribosomes. The kinase MAPKAPK2 is known to be involved in mRNA stabilisation.

Asthagiri et al. [2000] has shown that during cell cycle progression the integrated response I of certain downstream kinases (ERK kinase) provide an appropriate metric for quantifying the dependence of DNA synthesis on the degree of the activation of these kinases. Cell proliferation and DNA synthesis must coincide with cell growth, and therefore with the efficiency of the translation machinery. Thus, the integrated responses I of certain kinases might be of particular interest for a proper protein translation process. For the network under consideration, those kinases with high values for integrated response also possess high amplitude values (data not shown). The distribution of amplitudes and integrals of kinases give reason to ask if there is a distinction of kinases with respect to their activation strengths. Compared to "upstream" kinases which may act solely as information transmitters, the total activation strengths of the "downstream" kinases involved in the translation process (p70S6K, MAPKAPK2, Mnk, MSK) are likely to be of special importance.

Distributions of the three dynamic characteristics are depicted in Figure 2.24. The upper panel shows the distribution of amplitudes which is quite non-uniform. As one can already see in Figure 2.23, most kinases have small amplitudes and only very few exhibit high signal amplification with $S > 10$. Both signalling times and durations are more dispersed. The high peak for small values of all three characteristics is due to the activation profiles of the input kinases which are exactly the same. Values for τ and ϑ exhibit a high correlation with a correlation coefficient of $R^2_{\tau,\vartheta} \approx 0.98$. Amplitudes of kinases and signalling times or durations show much lower correlation ($R^2_{S,\tau} \approx 0.6$, $R^2_{S,\vartheta} \approx 0.6$). These calculations have been performed with a specific set of parameters ($\beta/\alpha = \lambda = 1.0$). However, other values would merely shift the time scales, thus maintaining qualitatively the distributions and correlation patterns.

To assess the dynamic characteristics of the transpath network more thoroughly, 100 randomised networks are generated for a direct comparison. These random networks are not allowed to contain cycles which ensures that they possess stable off-states. Thus, for the dynamic simulations the same parameters can be used as for simulations of the transpath network. Moreover, the same 25 input kinases can be deployed for receptor attachments, because the algorithm for generating random networks does not change node degrees and thereby keeps the number of input kinases of the initial TRANSPATH

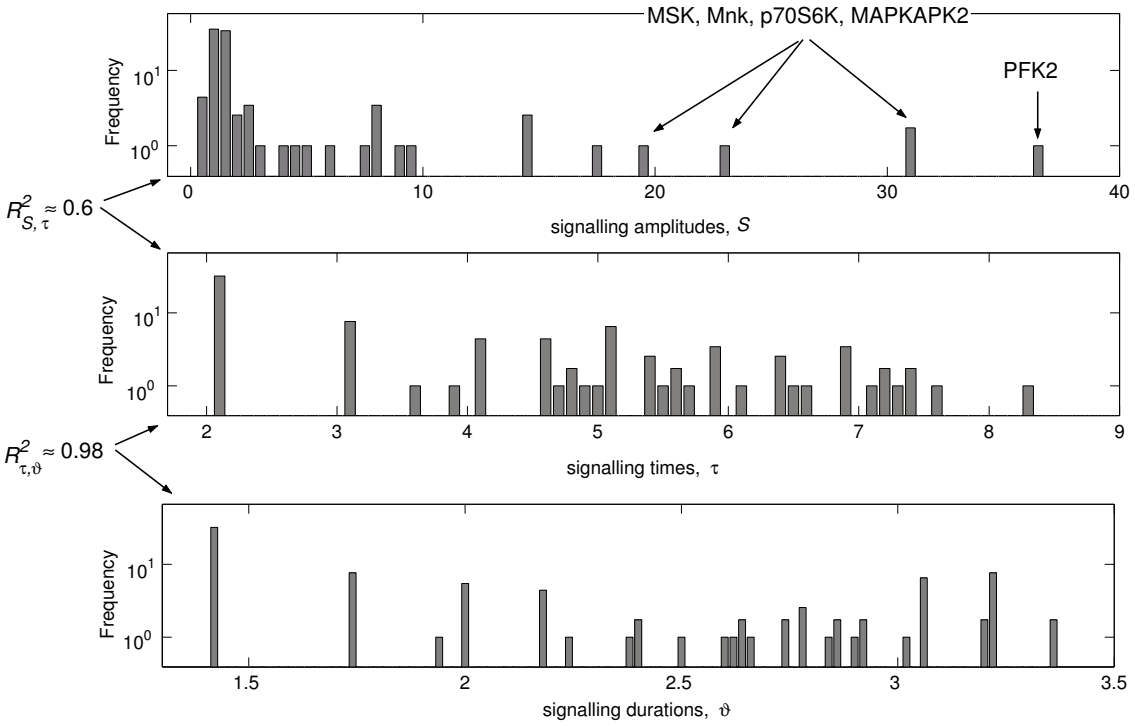


Figure 2.24: Dynamic properties of the kinase network from TRANSPATH. Distribution of kinase amplitudes S_i (A), signalling times τ_i (B), and signalling durations ϑ_i (C). The correlation coefficient between amplitudes S_i and signalling times τ_i is $R^2_{S,\tau} \approx 0.6$. A similar value is obtained for the correlation between amplitudes and signalling durations of the kinases ($R^2_{S,\vartheta} \approx 0.6$). In contrast to that, the correlation between the values of τ_i and ϑ_i is higher: $R^2_{\tau,\vartheta} \approx 0.98$.

network.

Similar to the TRANSPATH network, the distributions of signalling times and durations for the random networks are also rather evenly distributed (data not shown). However, unlike the kinase amplitudes in the TRANSPATH network, the amplitudes of kinases in random networks are more uniformly distributed. Moreover, in each of these random networks, kinase amplitudes and signalling times or durations tend to be more correlated. To illustrate these differences with respect to the distributions of key values appropriately, the correlation coefficients $R^2_{S,\tau}$ (A), $R^2_{S,\vartheta}$ (B) and $R^2_{\tau,\vartheta}$ (C) for all 100 random networks are calculated and shown in Figure 2.25. The correlation coefficients $R^2_{S,\tau}$ and $R^2_{S,\vartheta}$ of random networks are rather normally distributed (Figure 2.25A and B). In both cases the smallest values are slightly smaller than 0.6. Interestingly, the val-

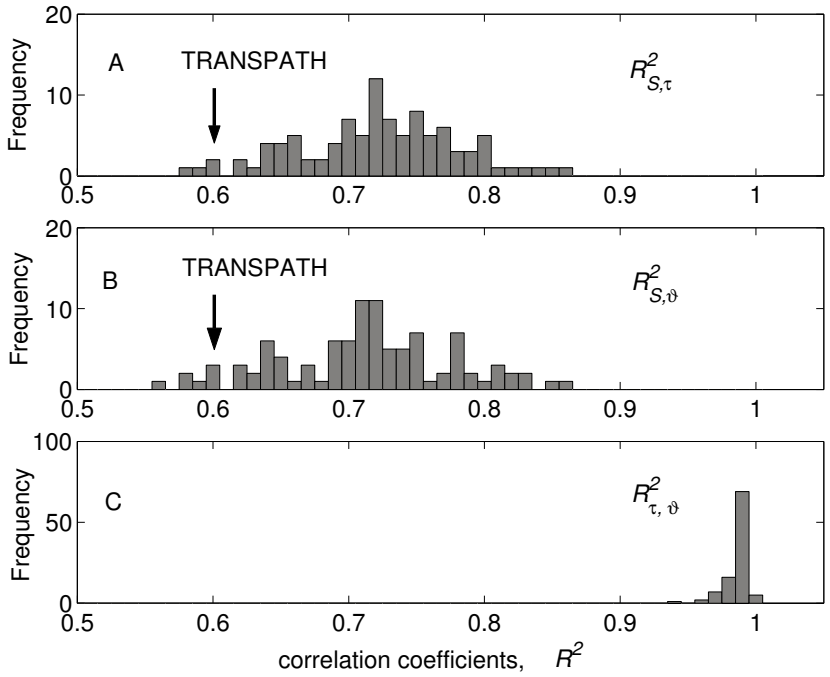


Figure 2.25: Dynamic properties of 100 randomised kinase networks. Correlation coefficients of kinase amplitudes and signalling times $R^2_{S,\tau}$ (A), kinase amplitudes and signalling durations $R^2_{S,\vartheta}$ (B) and signalling times and signalling durations $R^2_{\tau,\vartheta}$ (C).

ues of the two corresponding correlation coefficients of the TRANSPATH network (see legend of Figure 2.24) are at the lower limits of correlation coefficients which are apparently possible between kinase amplitudes and signalling times (or signalling durations) in networks of the same size.

Such a "decoupling" would make sense in the light of a kinase network evolution which selectively prefers those designs enabling long signalling durations (or signalling times) with moderate amplitudes (or vice versa) simultaneously.

Figure 2.25C depicts the distribution of correlation coefficients between signalling times and duration. For random networks as well as the TRANSPATH network these two characteristics are highly correlated.

2.3 Discussion

Although signalling pathways are becoming increasingly a subject of mathematic modelling, the explanation of the design of these systems is still an unsolved problem. This

might be partly due to the fact that – in contrast to metabolic networks – signalling networks consist of a higher variety of distinct components (adaptor proteins, scaffolds, several kinds of G-proteins, kinases and phosphatases) and regulatory mechanisms. The present model is an extension of previous investigations by Heinrich et al. [2002].

In contrast to simulation models which refer to specific pathways, large sets of networks with different structures and size are studied. This allows for identifying groups of network designs with specific dynamic properties such as signal amplification or dynamic stability.

The family of networks with four kinases is investigated analytically with respect to their structural robustness. Therefore, one considers the fraction of stable and amplifying networks which can emerge after a mutation of an initial (stable and amplifying) network. Specific values for β/α can be figured out which yield the most robust scenarios indicated by a maximum fraction of networks remaining stable (weak robustness) or stable and moreover amplifying (strong robustness) after a mutation. For future projects, it is planned to extend these calculations to larger networks, and as a long-term strategy it might also be interesting to compare β/α-values derived from this type of studies to values from the literature.

Based on the analysis with small networks, predictions are made concerning an optimal design of larger signalling networks in terms of network connectivity and number of cycles. The kinase network retrieved from the database TRANSPATH has indeed a very low connectivity, and no cycles are found. Calculations with random networks (null hypothesis) show that the probability of finding acyclic networks at random is very low ($P_{\mathrm{DAG}} \approx 0.04$). Further empirical studies on the structure of the kinase network from TRANSPATH reveal other serious deviations from a random structure, demonstrating that this system can be regarded as the outcome of mutation and selection processes.

In the TRANSPATH network, the sets of routes between input and output kinases are composed mostly of pathways of the same (shortest) length. In contrast to that, the set of routes between input and output kinases in random networks consist mostly of pathways of multiple lengths.

A measure η evaluates the crosstalk strength between all pathway pairs which do not contain a common kinase (disjoint routes). Two pathways having a small η value possess only few cross interactions compared to the maximum cross interactions which would in principle be possible. In the TRANSPATH network, a relatively high number of pathways do not show any, and many exhibit only little crosstalk. There is a rather monotonous decrease of the number of pathway pairs displaying higher η values. The picture is different for random networks, where the frequencies of η values are rather

normally distributed. A possible extension of this type of crosstalk studies might be to consider more closely the differences in the amount of crosstalk within consecutive layers of cascades (see Figure 2.21). Signalling pathways defined differently from the approach used here may also be considered, e.g. pathways which are distinguished from each other by additional biological information.

Scale-freeness and the small world property refer to a specific distribution of connectivities of components in metabolism. These properties are often discussed in conne-xion with the property of robustness and vulnerability of metabolic networks. Similarly, one may ask for reasons of the specific distribution of kinase connectivities in signalling networks. Interestingly, random networks with the same node degrees as the TRANSPATH network, generated with additional constraints concerning small cycles, show a tendency to avoid cycles entirely. Assuming that only properly arranged kinases lead to a pathway activation, scaffold proteins may also be regarded as a mechanism to avoid small cycles, i.e. that downstream kinases activate upstream kinases. It will be worth analysing, whether in other networks, e.g. gene regulatory networks, cycles may emerge more easily by structural mutations, since they might play an important role therein.

Dynamic properties of the design of the kinase network are studied by carrying out simulations with ODEs. In contrast to random networks, only few kinases have significantly high amplitudes (kinases involved in protein translation). Another distinguishing feature is that amplitudes are less correlated to signal durations than in the case of random networks – a characteristic indicating that the design is appropriate for a decoupling of (high) signalling amplitudes and (long) signalling durations.

Low connectivity, low number of cycles, a characteristic composition of pathways and a characteristic distribution of crosstalk intensities may be considered as design properties of the TRANSPATH network which are the outcome of an evolutionary selection process. A remaining task to corroborate these results is to update the kinase network by taking into account more recent information accumulated in the TRANSPATH database as well as information of other databases. This may further increase the degree of reliance concerning the wiring of the network.

Chapter 3

Control mechanisms of signal transduction networks

Signal transduction is of crucial importance to cells to reach cellular decisions with respect to division, differentiation and apoptosis. The complexity of signal transduction networks is overwhelming because of the large numbers of interacting constituents, and because of their complicated circuitry, involving feed forward, negative feedback and crosstalk, as well as the fact that the constituents are highly diverse (kinases, phosphatases, adaptors, G-proteins etc.). Due to this complexity, apparently simple but highly important questions remain hitherto unsolved. For instance, is a small number of reactions responsible for the behaviour of the entire system (and ultimately, for cellular behaviour) and if so, which reactions are they? Is the control of signal transduction and hence of proliferation, differentiation and apoptosis always evenly distributed within the network?

Such questions may be answered by detailed and extensive quantitative experiments with inhibitors and activators of signal transduction proteins. However, the arsenal and the specificity of these are limited. Mathematical modelling can assist in such cases. Previously, modelling efforts have significantly enhanced the understanding of control mechanisms of signal transduction pathways (Heinrich et al. [2002]), Lee et al. [2003]).

It is known that mutations may result in a constitutive activation of the mitogen-activated protein kinase (MAPK) pathway, for example in many human tumour cells (Hoshino et al. [1999]). The analysis of which proteins and reactions are in control of a signalling pathway is of vital importance not only for understanding of the system, but also for drug design, since it helps choosing potential drug targets according to the magnitude of their control on cell proliferation. Decisive for the type of cellular response

that is evoked by MAPK signalling are the magnitude and duration (transient versus sustained) of ERK activation (Cook et al. [1999]; Marshall [1995]; Tombes et al. [1998]). Therefore, it is of importance to understand which processes and proteins determine the dynamic time profile of ERK activity and hence the choice between proliferation and differentiation.

Metabolic Control Analysis (MCA) was developed with steady states situations in mind (Heinrich and Rapoport [1974]; Kacser and Burns [1973]) but for signal transduction steady state levels may be less relevant and the important features are the amplitude of a transient peak, the duration of signalling and the integrated response. In this chapter, MCA is applied to investigate these dynamic properties considering various models of signalling systems.

In a first step, the control properties of kinases and phosphatases towards the time profile of active ERK are investigated using kinase and phosphatase inhibitors. Experimental data from Hornberg, J. and colleagues[1] are used for a direct comparison of a simple model of signalling cascades with the results from these inhibitor experiments. The model is employed to simulate different activating and inhibitory interventions of the involved reaction rates. This analysis is based on and confirms results from theoretical studies presented in Heinrich et al. [2002].

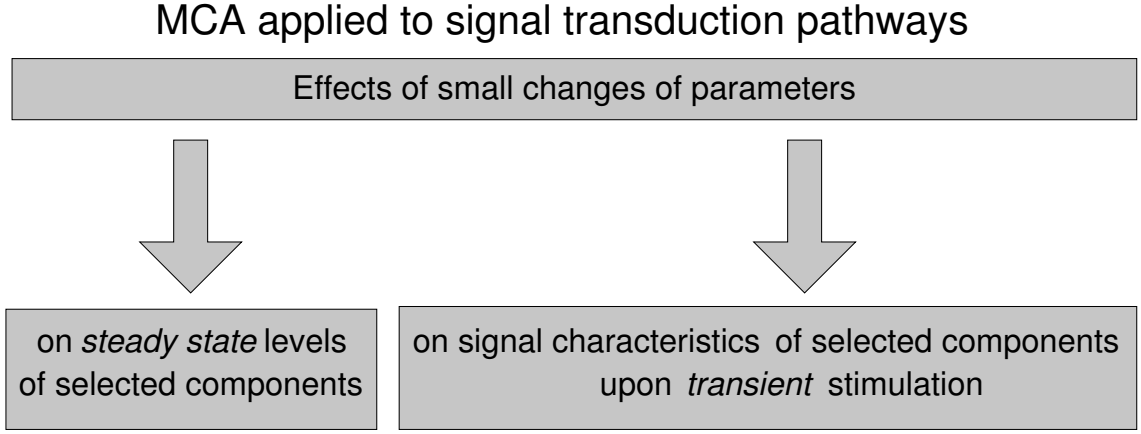
In a second step, the method is applied to a specific model of EGF signalling published by Schoeberl et al. [2002]. The importance of each reaction in this complex network is determined with respect to its biologically relevant output, the transient phosphorylation profile of ERK. It is shown that only a small proportion of reactions within the network exhibit control. Control coefficients of key values like signal amplitude, signal duration and the integrated response take the central role in this analysis. Effects of mutations and changes in gene expression are also examined by calculating response coefficients of reasonable enzyme concentrations.

Last, the question is addressed if there is an optimal design of signalling cascades with respect to the sensitivity of the signal amplitude. Sensitivity is defined within the framework of MCA by considering the standard deviation of all control coefficients for the signal amplitude. Taking this measure as an objective function to be maximised, cascades of different length n and phosphorylation mechanisms (mono/dual) are evaluated. This more general approach is inspired by investigations of the "zero-order ultrasensitivity" of signalling cascades (Goldbeter and Koshland [1981]; Goldbeter and Koshland [1982]; Huang and Ferrell [1996]).

[1]Department of Molecular Cell Physiology, Institute of Molecular Cell Biology, Vrije Universiteit Amsterdam

3.1 Control coefficients and summation theorems for transient activation profiles

Metabolic control analysis (MCA) is a phenomenological sensitivity analysis which can be used to describe quantitatively how enzymes in a metabolic pathway control fluxes and metabolite concentrations. In MCA one studies the relative control exerted by each step (enzyme) on the system's variables. By applying a perturbation to the step being studied the effect on the variable of interest is measured after the system has settled to a new steady state. MCA has been extensively applied to metabolism where mainly steady states are of interest. In signal transduction steady state concentrations of selected components have also been successfully considered (Lee et al. [2003]; Kholodenko et al. [1997]). However, the important operating regime of signalling pathways is different; here, signal amplitude, signal duration and the integrated response are relevant characteristics (see Figure 3.1). The integrated response I of a time profile of a compo-



$$C_i^{X_j} = \frac{\partial \ln X_j}{\partial \ln v_i} = \frac{k_i}{X_j} \cdot \frac{\partial X_j}{\partial k_i} \quad ; \quad C_i^{\{A,d,I\}} = \frac{\partial \ln\{A,d,I\}}{\partial \ln v_i} = \frac{k_i}{\{A,d,I\}} \cdot \frac{\partial\{A,d,I\}}{\partial k_i}$$

Figure 3.1: Control theory in signal transduction. Under steady state condition the control of a reaction v_i on the concentration of a selected component X_j is of interest. Upon transient stimulation well defined features like signal amplitude A, signal duration d and integrated response I of the selected component X_j are important. For reaction rates depending linearly on rate constants k_i, the second equalities hold: $v_i \propto k_i \Rightarrow \partial \ln v_i = \partial \ln k_i$.

nent $X_j(t)$ is defined in Eq. (2.15), where – as an specific example – the concentration of a phosphorylated kinase is considered. Signal amplification and signal duration can be

defined in several ways. In chapter 2 definitions of these characteristics are used which are adequate for analysing large sets of networks designs analytically. For the numerical calculations of control coefficients in this chapter it is more convenient to choose definitions different from Eq. (2.13) and Eq. (2.15). The amplitude of an activation profile is taken as the maximum value that the profile attains and the duration is defined as the total time for which $X_j(t)$ exceeds a certain fixed value. For example in the cascade model of section 3.2 this value is chosen as 5% of the total concentration of the ERK kinase and in section 3.3 the duration is quantified at a level of 10% of the amplitude of the ERK activation profile. To indicate the difference to the previous definitions, the letters A and d are chosen for amplitude and duration, respectively. Figure 3.7 illustrates the employed definitions for these quantities.

As part of the theory, MCA embodies summation theorems for control coefficients. Summation theorems relate the sum of the all control coefficients of a certain systemic property to a constant value. In Appendix B.1 summation laws for the control of time dependent phenomena are given. For the control coefficients of the time dependent profiles $X_j(t)$ at a certain time point $t = t_0$ one derives:

$$\sum_i C_i^{X_j} - \frac{\partial \ln X_j(t)}{\partial \ln t}\Big|_{t=t_0} = 0 \ , \tag{3.1}$$

where the second term is the (scaled) derivative of $X_j(t)$ at time $t = t_0$. This term is zero for the maximum of the time profile $X_j(t)$, i.e. for $X_j(t_0) = A$. For the control coefficients of A, d and I of $X_j(t)$ one arrives at the following summation theorems:

$$A: \quad \sum_i C_i^A = 0 \quad ; \quad d: \quad \sum_i C_i^d = -1 \quad ; \quad I: \quad \sum_i C_i^I = -1. \tag{3.2}$$

It is, however, worth mentioning that the summation laws are not restricted to these specific definitions for amplitude, duration and integrated response. In particular they are also valid for the quantities defined in Eqs. (2.13–2.15) (see Appendix B.1).

The summation laws dictate that the control coefficients for the rates v_i of all processes i add up to constant values, e.g. the sum of all control coefficients on the amplitude equals 0. This implies that all reactions of the network with positive control coefficients, taken together, are exactly as important for the amplitude as all reactions with negative control, i.e. each of these two groups have the same quantitative impact on the amplitude. All control coefficients on the duration of signalling and the integrated response must add up to -1. This implies that the inactivating reactions which cause shorter duration and therefore have negative control coefficients exert more control on these properties than activating reactions which cause longer duration and therefore

have positive control coefficients. For simple linear cascades with kinases as activating and phosphatases as inactivating processes, the basis of these control principles was analytically tracked down in Heinrich et al. [2002].

In section 2.1.5 the structural robustness of signalling networks is analysed. Within the framework of MCA one can define a measure for robustness and sensitivity of a signalling property based on changes of kinetic parameters (Lee et al. [2003]). The larger the control coefficients, the higher the corresponding sensitivity of a signalling property becomes. As an appropriate metric the standard deviation of control coefficients from their mean value is considered as a measure for the sensitivity. In the following analysis, the focus lies on the sensitivity of the signal amplitude. According to the summation theorem the mean value of control coefficients is zero in this case. Therefore the overall sensitivity for a signal amplitude is given as follows:

$$\sigma^A = \sqrt{\frac{1}{n}\sum_i (C_i^A)^2}, \tag{3.3}$$

where the summation is performed over all reaction steps. High values of σ^A indicate that the given amplitude is on average very sensitive with respect to changes of rate constants. Based on the sensitivity, one can easily define a measure for robustness ρ^A:

$$\rho^A = \frac{1}{1+\sigma^A}. \tag{3.4}$$

As σ^A may vary between zero and infinity, the range of ρ^A is confined to the interval $1 \geq \rho^A \geq 0$. High values of ρ^A resulting from low σ^A values indicate that the amplitude is robust against parameter perturbations. Other definitions of robustness via control coefficients might be applicable as well, e.g. the function $\rho^A = \exp(-\sigma^A)$ also takes values confined to the interval $1 \geq \rho^A \geq 0$.

In the next three sections control mechanisms of signalling pathways are studied by (1) reproducing the results of inhibitor experiments aimed at kinases and phosphatases with simulated inhibitions of these components in a simple kinase cascade model, (2) applying the concept of "transient" MCA to a comprehensive computational model of EGF signalling, (3) evaluating a possible optimal length of cascades with respect to a high efficiency of controlling the amplitude of the signal output (overall sensitivity).

3.2 Control principles of phosphatases and kinases: modelling inhibitor experiments

General control principles that govern signalling cascades can be validated by examining effects of kinase and phosphatase inhibition on a MAP kinase pathway. In this section inhibitor experiments and a simple model which mimic the results are described. Quantitative measures for the control of signal amplitude, duration, integrated response and the final signal strength (final steady state level) confirm the principles which are reflected in the summation theorems for control coefficients (see Eq. 3.2): Total control on signal amplitude and on final signal strength amounts to zero, and total control on signal duration and on integrated response is -1. Therefore, there is a tendency that the impact of processes with positive control (e.g. kinases) on duration and integrated response remains lower compared to processes with negative control (e.g. phosphatases).

These principles which are studied analytically in Heinrich et al. [2002] are illustrated by a direct comparison of experimental results with a simple in silico model described in the next section.

For the experiments rat kidney (NRK) fibroblasts are used, because the cell cycle of these cells can be synchronised relatively easily, causing all cells to behave similarly in response to external stimuli. They are used frequently as a model system to study cellular alterations that accompany oncogenic transformation (van Zoelen [1991]). Activation of the ERK pathway is required for the proliferation of fibroblasts (Pagès et al. [1993]). The pathway under consideration consists of three kinases in succession (Raf, MEK and ERK) and can be activated by various extracellular stimuli, including the epidermal growth factor (EGF). The experiments were carried out by Hornberg, J. and colleagues at the Department of Molecular Cell Physiology, Institute of Molecular Cell Biology, Vrije Universiteit Amsterdam.

3.2.1 Model description

The mathematical model consists of a receptor and three consecutive kinase/phosphatase monocycles (Figure 3.2A).

The receptor is activated instantaneously at time $t = 0$ by added EGF and then inactivated over time (R_i). The inactive form of the receptor is re-circulated slowly to become active once again because EGF remains present. The active form of the receptor phosphorylates and thereby activates the first kinase to become X_1^*. Through phosphorylation this kinase then activates the second kinase, which in turn activates a

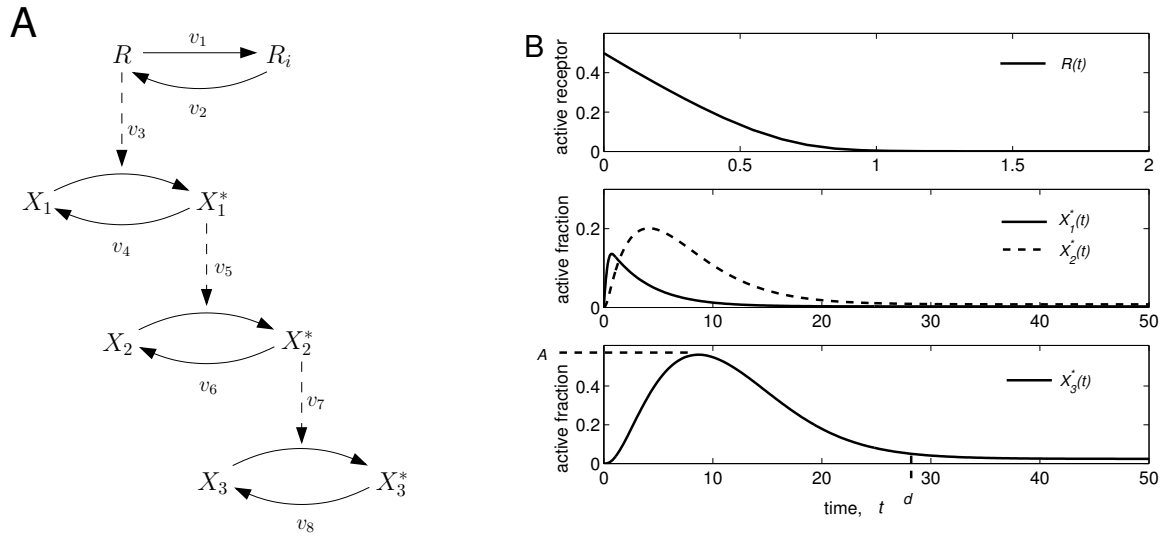


Figure 3.2: Schematic model representation and time profiles of the active model components. A) Active receptor with concentration R activates a cascade of three kinase/phosphatase monocycles. The ligand remains present and, therefore, the inactive receptor is taken to slowly recycle, in order to become active again (v_2). B) Time profiles of the four active components. The receptor is inactivated relatively quickly to attain a very low steady state concentration. The active kinases reach maximal peak concentrations and then decrease to low steady state concentrations. Definitions of amplitude A, duration d and integrated response are given in the text.

third kinase.

All reaction steps follow Michaelis-Menten kinetic rate equations, with $V_{\max} = 1.0$ and $K_{\mathrm{m}} = 0.1$ for the activating reactions (v_3, v_5, v_7) and the receptor inactivation reaction (v_1); $V_{\max} = 0.3$ and $K_{\mathrm{m}} = 1.0$ for the deactivating phosphatase reactions (v_4, v_6, v_8); and $V_{max} = 0.01$ and $K_{\mathrm{m}} = 0.1$ for the receptor re-circulation reaction. The initial concentrations for R_i and the active kinases X_1^*, X_2^* and X_3^* are set to 0 and for the active receptor R to 0.5. The initial concentrations for the inactive kinases X_1, X_2 and X_3 are chosen to be 1. All concentrations are in mM. The kinase reactions depend proportionally on the concentration of their corresponding activators, i.e. the preceding phosphorylated kinases. The model and the specific parameters are chosen not to accurately describe the MAPK pathway but to analyse control principles of signalling cascades by comparing the model behaviour due to perturbations of reactions with the according performance of inhibitor experiments. The model equations are as follows (for

details see Appendix B.4):

$$\dot{R} = v_2 - v_1 = \frac{0.01 \cdot R_i}{0.1 + R_i} - \frac{R}{0.1 + R} \tag{3.5}$$

$$\dot{X}_1^* = v_3 - v_4 = \frac{R \cdot X_1}{0.1 + X_1} - \frac{0.3 \cdot X_1^*}{1 + X_1^*} \tag{3.6}$$

$$\dot{X}_2^* = v_5 - v_6 = \frac{X_1^* \cdot X_2}{0.1 + X_2} - \frac{0.3 \cdot X_2^*}{0.1 + X_2^*} \tag{3.7}$$

$$\dot{X}_3^* = v_7 - v_8 = \frac{X_2^* \cdot X_3}{0.1 + X_3} - \frac{0.3 \cdot X_3^*}{1 + X_3^*} \tag{3.8}$$

In Figure 3.2B the activation profiles for the active components are shown. Whilst the concentration of the active receptor R declines over time, the three consecutive kinases are activated, i.e. phosphorylated, reach a peak value and subsequently decline to levels that exceeds the level before receptor activation. These time-patterns for activation of the components of the MAPK cascade are commensurate with what is reported experimentally for many cell types and with the experimental results which will be presented here. Therefore, this model is used to examine how these time patterns are controlled by the kinases and the phosphatases.

For a later quantification of control via control coefficients, signalling characteristics are defined similarly to section 3.3: (i) signalling amplitude A, which is taken as the maximal X_3^* concentration; (ii) duration of signalling d that is defined as the time point at which the X_3^* concentration drops below an arbitrarily chosen value of 5% of the total concentration of the third kinase (= 0.05); (iii) the "integrated response", i.e. the integral of the X_3^* time profile until this duration time point d; (iv) the final signalling level, i.e. the steady state of X_3^*.

In order to examine how the second kinase in the cascade determines the time dependence of the activity of the third kinase, only $V_{\max}$ of the second kinase reaction (v_5)is varied stepwise and for each value of $V_{\max}$ the system is recalculated. This modulation corresponds to the experiment to be described below in which MEK, the second kinase of the MAPK pathway, will be inhibited by the non-competitive inhibitor PD98059 (Alessi et al. [1995]). The effect of such an inhibitor is studied by decreasing the $V_{\max}$-value from 1.0 (kinase inhibition 0%) to 0.4 (kinase inhibition 60%) by 20% intervals. The result is shown in Figure 3.3A. The duration of the peak also decreases, but much less so; e.g. an inhibition that decreases the amplitude by 25%, advances the time at which the signal returns to below 0.05 by about 10%. The final level of X_3^* decreases substantially when calculated in relative terms: The final level is already low before the kinase is inhibited. The effect of signal transduction on transcription of downstream genes might be less a

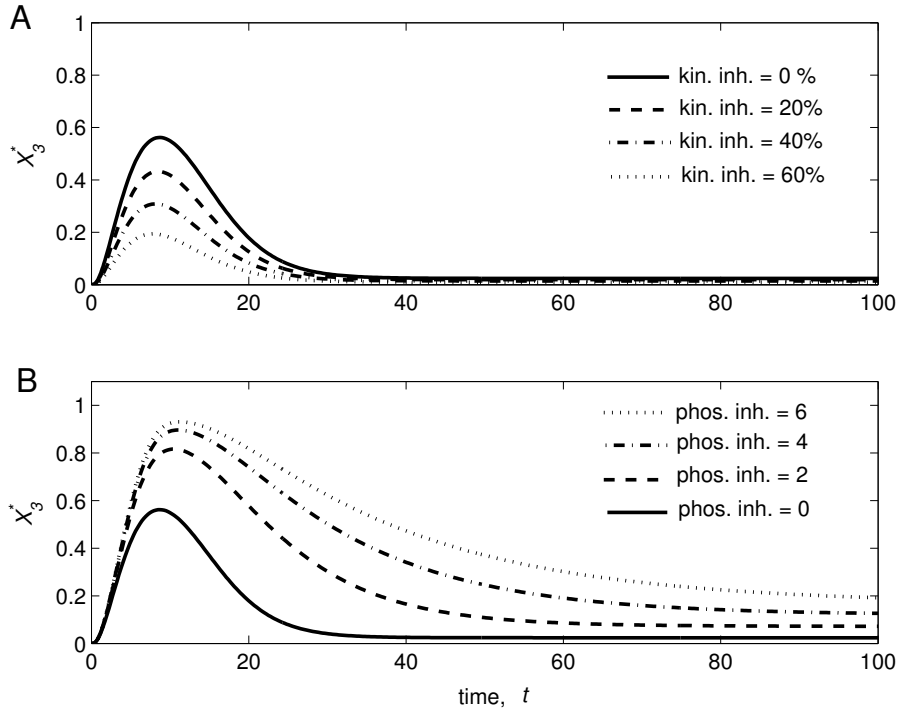


Figure 3.3: Effect of kinase and phosphatase inhibition in the X_3^* time profile. A: The second kinase reaction in the model (v_5) is increasingly inhibited, having a large decreasing effect on the amplitude and a less pronounced effect on duration. B: Increasing inhibition of the third phosphatase (v_8) which causes an increase in both amplitude and duration of signalling. Apparently, both, kinase and phosphatase inhibition, also affect the integrated response and steady state X_3^* concentration.

function of the amplitude of X_3^* than of the integrated concentration of this factor. For the area under the curve (integrated response) one finds that kinase inhibition has a considerable effect on this, too.

Essentially, it can be seen that in this example model, the activity of the second kinase exerts substantial control on signalling amplitude, both, in the initial phase of signalling and much later, and exerts less control on the duration of signalling.

To examine the influence of phosphatases on signalling kinetics an inhibitor with concentration Y is introduced. The inhibitor is supposed to competitively inhibit the dephosphorylation of X_3^*. In this way an experiment is anticipated (see below) in which protein tyrosine phosphatases are inhibited. For such an inhibitor the rate equation v_8 in Eq. (3.8) changes as follows: $v_8 \rightarrow v_8 = \dfrac{0.3 \cdot X_3^*}{1 + X_3^* + Y}$. To demonstrate the effect of

inhibition of the third phosphatase in the model, the concentration Y is increased from 0 to 2, 4 and 6, respectively.

With increasing inhibitor concentration, the X_3^* amplitude becomes substantially higher (Figure 3.3B). In addition, the inhibitor increases the duration of the peak dramatically, prolonging X_3^* signalling. For instance an inhibitor concentration that increases the peak height by one third, doubles the time required for the X_3^* concentration to drop below (an arbitrarily chosen value) 0.1. As can be seen in Figure 3.3B phosphatase inhibition also increases the final level of X_3^* and the integrated response quite substantially.

For this example model, the calculations show that phosphatases and kinases are equally important for two characteristics of signal transduction, the peak amplitude and the final activation level, whereas the duration of signalling, and the integrated response might be more the control domain of the phosphatases exclusively.

For a further quantification of control exerted by the individual kinases and phosphatases the concept of control coefficients can be used. Normalised control coefficients, calculated with 1% perturbations, are shown in the Table B.5 in the Appendix B.4.

3.2.2 Experimental validation of control principles

For the stimulation experiments the NRK fibroblast cells are grown in culture dishes to sub-confluency and then serum-starved for three days in order to be arrested in the G_0-phase of the cell cycle. Subsequently cells are stimulated with 10 ng/ml EGF for the indicated periods of time. Where indicated MEK is inhibited by preincubation for 1 hour with various concentrations of the non-competitive inhibitor PD98059 (Alessi et al. [1995]). For the phosphatase inhibition (PTPs) cells are pre-incubated for 1 hour with 0.20 mM sodium orthovanadate (Huyer et al. [1997]). Western blot analalysis is used to quantify the time profiles of the kinase activations.

Stimulation with EGF is carried out in the presence of various concentrations of the non-competitive MEK inhibitor PD98059. A biphasic ERK-PP time profile can be observed, consisting of a rapid high peak followed by a low quasi-steady state (Figure 3.4A). Increasing MEK inhibitor concentrations results in decreased peak heights. Adding an inhibitor do not much affect the duration (width) of the peak. These results confirm the model predictions that kinase inhibition affects signalling amplitude much more than signalling duration.

In the model simulation, phosphatases control the signalling amplitude, signalling duration, final signal level and integrated response. To test these control principles ex-

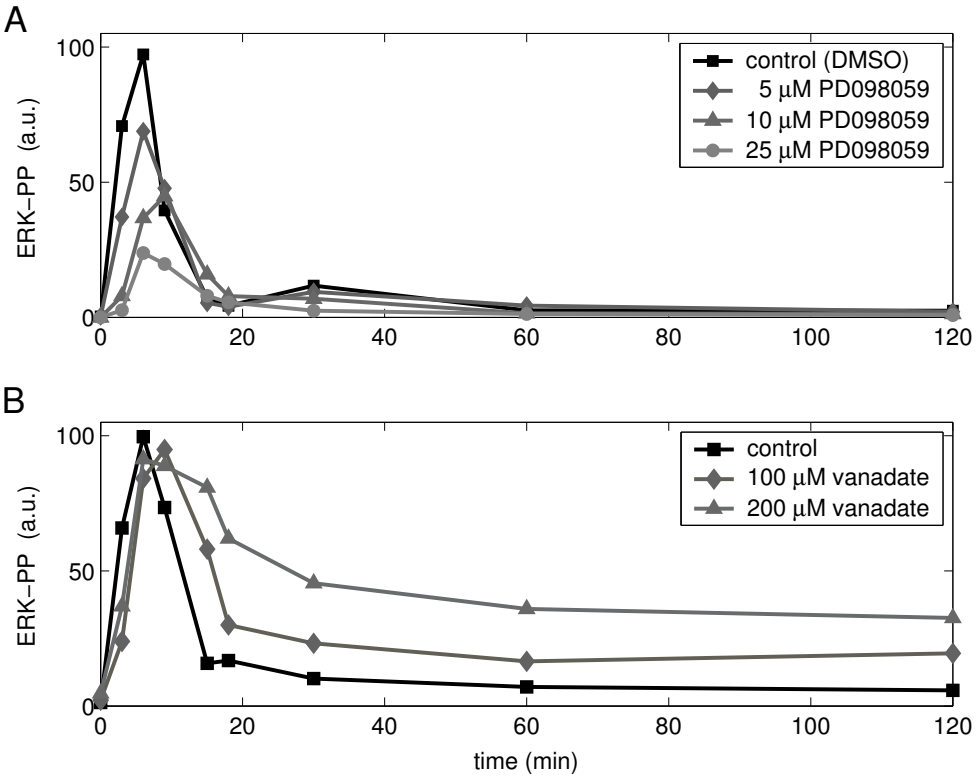


Figure 3.4: Experimental validation of the theoretical results. A: Inhibition of MEK, the second kinase in the MAPK pathway to ERK, using the non-competitive inhibitor PD098059, leads to a decrease in the peak ERK-PP concentration. B: Protein tyrosine phosphatase inhibition, using orthovanadate, significantly increases the duration of the ERK-PP activation profile.

perimentally, the protein tyrosine phosphatase (PTP) inhibitor sodium orthovanadate is applied to the cells which are subsequently stimulated with EGF. PTP inhibition results in a broader peak followed by a relatively high final quasi-steady ERK-PP concentration (Figure 3.4B), both in consonance with the model predictions.

The amplitude in the presence of the phosphatase inhibitor is no higher than that in control cells. This result corresponds closely to high kinase activations leading to saturation effects of the ERK-PP activation profile. A higher agreement with the outcome of the inhibitor experiments (saturation effects) can be achieved by using higher activities for all kinases in the model. Thefore, the model simulation is repeated with increased reaction rates for the kinases: The $V_{\max}$-values of the kinase reactions are set to 1.2. These higher kinase activations cause virtually all X_3 to be activated. The result

is depicted in Figure 3.5. Inhibition of the second kinase reaction (v_5) again decreased

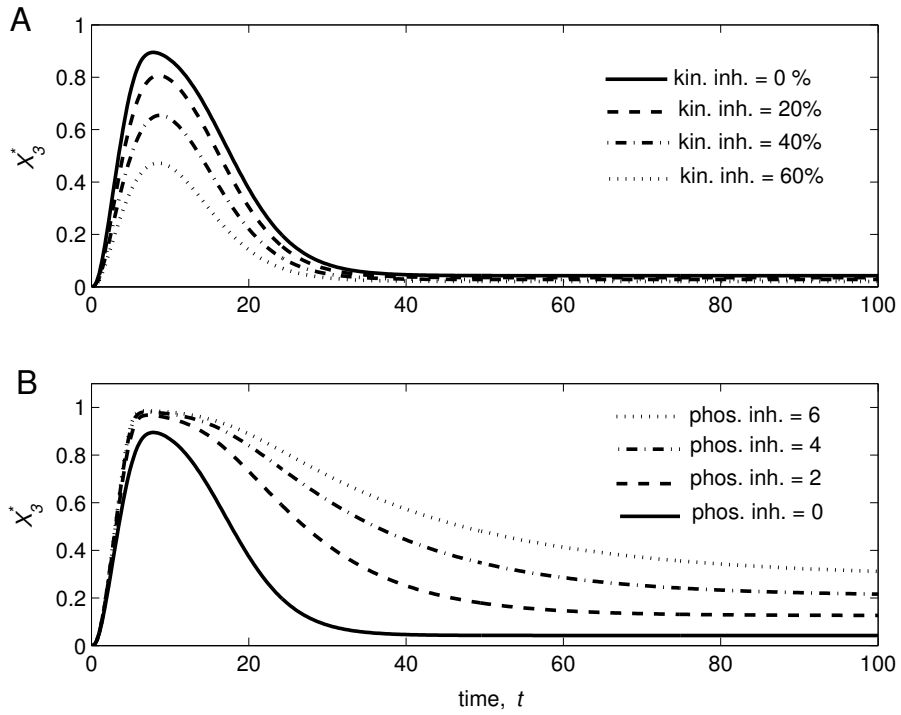


Figure 3.5: Simulations for different magnitudes of the kinetic parameters ($V_{\max} = 1.2$) for *all* kinases. The activation reactions are more active, leading to a higher peak X_3^* concentration.

the amplitude of signalling (A). Inhibition of the third phosphatase reaction (v_8) had some effect on the amplitude (B), but this is significantly smaller than in the case of Figure 3.3, since now virtually all X_3 is phosphorylated in the peak. Indeed, it has previously been shown that in NRK cells, this EGF concentration causes virtually all ERK to become doubly phosphorylated (Lahaye et al. [1998]).

The results of the present section show that general control principles that govern signalling cascades can be tracked down with inhibitor experiments and can moreover be described with a simple model. These findings correspond closely with those reported previously (Heinrich et al. [2002]). The control principles – different extent of control of kinases and phosphatases towards the ampltiude and duration of the signal output – are reflected in the summation theorems for the control coefficients.

The used model of a MAPK pathway is highly simplified. Many important components (e.g. adaptor and scaffold proteins) and processes (e.g. receptor internalisation) which are known to play a role in signalling are neglected. In the next section a more

elaborated model of the EGF-induced MAPK signalling pathway is studied with respect to the distribution of control among the processes involved in the cascade.

3.3 Control coefficients for a large EGF signalling model

3.3.1 Model description

The model published by Schoeberl et al. [2002] describes the dynamics of the mitogen-activated protein kinase (MAPK) cascade activated by the epidermal growth factor (EGF) receptor. The network is depicted in Figure 3.6. Signalling through the network is initiated by the binding of epidermal growth factor (EGF) to its receptor (EGFR). The signal is subsequently propagated by a series of events such as the recruitment of adaptor proteins, G-protein activation, and (de-)phosphorylation reactions, which ultimately lead to the activation of the extracellular signal-regulated kinase (ERK). This network stands at the root of many cellular processes, such as cell division (Cobb [1999]), and is constitutively activated in many human tumours (Hoshino et al. [1999]). Active ERK has several cytoplasmic and nuclear targets, such as transcription factors that influence expression of genes involved in cell cycle progression, and it is required for the proliferation of fibroblasts (Pagès et al. [1993]).

This detailed analysis of a specific signalling model asks whether control is as disperse as it can be in principle over the participating processes and components. The results are much less random than might expected and may contribute to explain oncogene distributions.

Since the model is continuously being improved and enlarged by the original authors, a detailed description of the version used here is provided in Appendix B.2. It is an updated version of previous models (Kholodenko et al. [1999]; Schoeberl et al. [2002]) and should be close to containing the best available experimental data on the pathway. The latest available version can be obtained from a website[2].

The computer replica of EGF-induced signalling through the MAPK cascade that is used here contains 148 molecular processes and 103 molecular species represented by system variables. It describes signal transduction all the way from EGF binding at the cellular membrane to the phosphorylation of ERK. The molecular processes include association, dissociation, as well as covalent (de–)modification and degradation events

[2]http://www.mpi-magdeburg.mpg.de/people/schoeberl/schoeberl.html

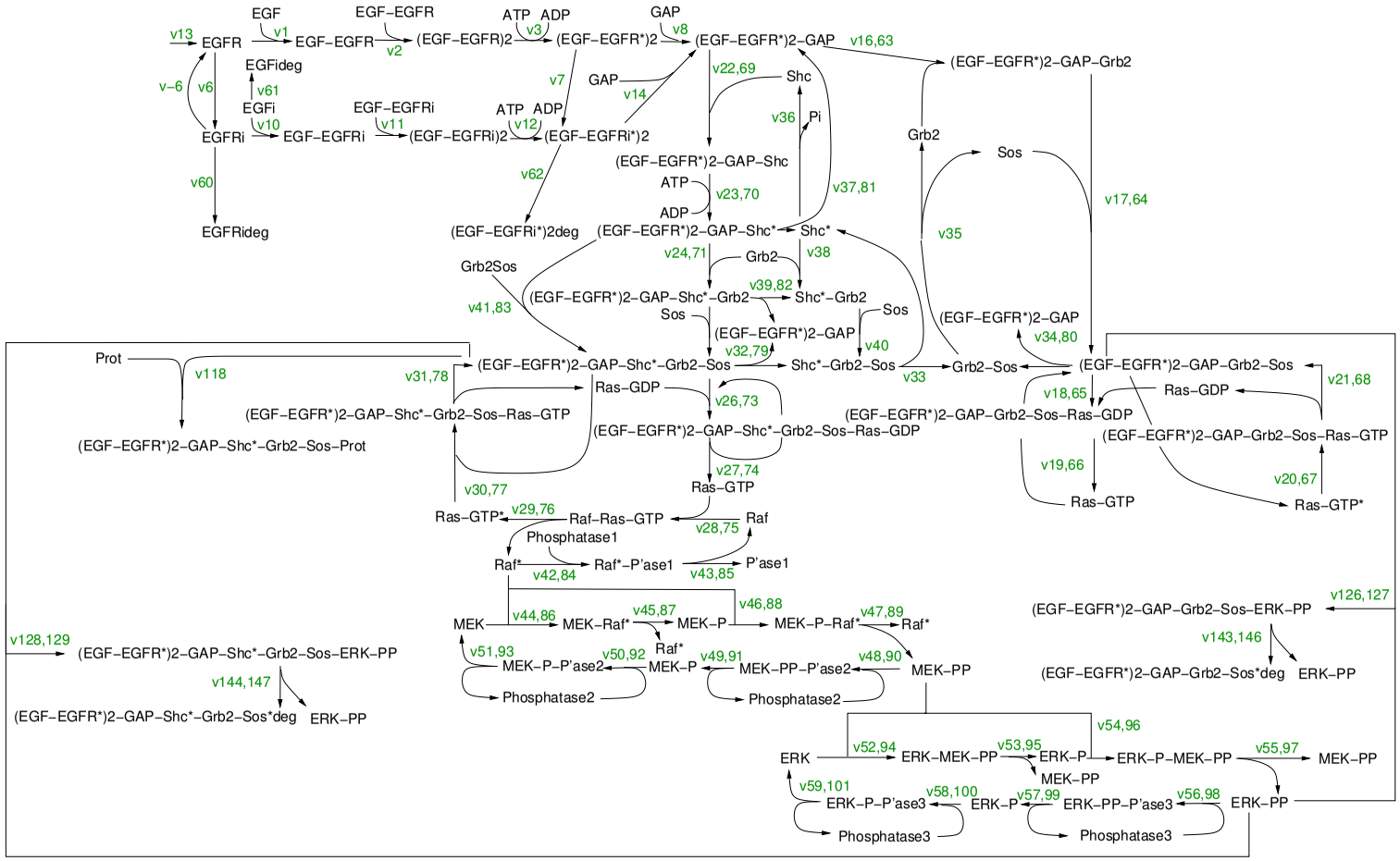


Figure 3.6: Biochemical reaction scheme of EGF receptor-induced MAP kinase cascade. The cascade can be initiated by Shc-dependent and independent pathways. The arrows represent the reactions specified in Appendix B.2 and are characterised by reaction rates v_i, $i = 1 \ldots 148$ (green numbers). The second green numbers identify reaction rates after internalisation. To simplify matters only selected reation rates are indicated.

and have all been described with mass-action kinetics.

Signalling is initiated by binding of EGF to EGFR, followed by the dimerisation and subsequent auto-phosphorylation of EGFR. Signal propagation involves either of two routes, both leading to the activation of the small G-protein Ras. One route starts with the binding of the "Src homology and collagen domain protein" (Shc) to the phosphorylated, ligand-bound, dimerised receptor followed by the binding of the growth factor receptor-binding protein 2 (Grb2). The other commences directly with the binding of Grb2. In both routes the association of Grb2 with the complex that includes EGFR recruits Sos to the cytoplasmic surface of the membrane. The recruitment of Sos brings the latter in close proximity to its membrane-anchored target Ras. Contact between Sos and Ras leads to the activation of the latter through the formation of Ras-GTP. Ras-GTP dissociates from the receptor complex, which then dissociates further. The inactivation of Ras-GTP involves the activity of a GTPase activating protein (GAP). Although Raf phosphorylation is subject to an ongoing discussion in the literature, the present model assumes that a free Ras-GTP molecule directly phosphorylates Raf. Phosphorylated Raf has kinase activity and does phosphorylate the kinase MEK which, in its doubly phosphorylated form, phosphorylates ERK. Raf, MEK and ERK are dephosphorylated by various phosphatases; in the model this is described in terms of the different phosphatase activities, one for each kinase. The model includes the negative feedback loop from doubly phosphorylated ERK (ERK-PP) to Sos, which causes the dissociation of Grb2-Sos from the receptor complex. Also included is an elaborate description of the processes involved in the internalisation and degradation of the various receptor complexes.

In the absence of stimulation by EGF the system is in a stable steady state, in which ERK is not phosphorylated (for other characteristics of this reference state, see Table B.2 in Appendix B.2). The model distinguishes between two pools of ERK-PP: one is located freely in the cytoplasm; the other is associated with the internalised receptor. A closer look reveals that this internalised ERK-PP makes up at most 1.2% of the total ERK-PP concentration. Figure 3.7 displays the computed transient profile of the total ERK-PP concentration, i.e. the sum of the two pools. Upon addition of EGF at $t = 0$ in a step function manner, the total concentration of ERK-PP rapidly increases to attain about 50% of the total concentration of ERK within about 5 minutes. Then, the ERK-PP concentration slowly decays to zero, passing 10% of its maximal activation level at about 65 minutes.

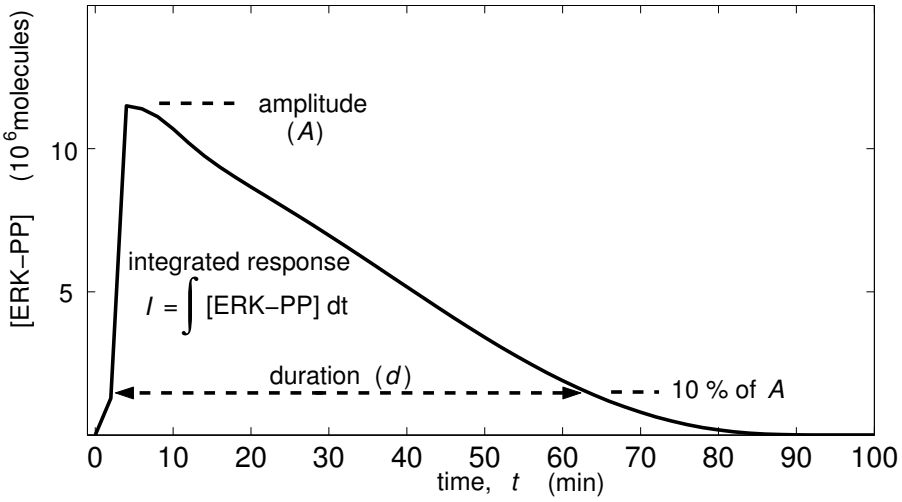


Figure 3.7: Transient activation profile of ERK-PP as function of time after the addition of EGF at time zero calculated with the kinetic model. The definitions of the three quantities that characterise the transient profile are depicted.

3.3.2 Results

The model is implemented in Mathematica 5.0. In order to calculate the control coefficients for the processes in the network, the reaction rates are perturbed by a small fractional change (numerically: $\partial k/k = 10^{-6}$). The forward and backward rate constants for the process are changed simultaneously to ascertain that the activity as a whole is perturbed with the same fractional change. In this way, the equilibrium constant remains fixed which is necessary to fulfil thermodynamic constraints. Control coefficients defined in MCA are calculated by infinitesimal perturbations. Since perturbations in living cells are generally larger than differentially small (e.g. as a result of a mutation), larger perturbations of the molecular activities are also applied ($\Delta k/k = \pm 0.5$).

The individual sums of the control coefficients (for small perturbations) on amplitude, duration and integrated response are in good accordance with the summation theorems of Eqs. (3.2):

$$\text{Amplitude of ERK-PP:} \qquad \sum_i C_i^A = 1.4 \cdot 10^{-6}.$$

$$\text{Duration of ERK-PP:} \qquad \sum_i C_i^d = -1.00008.$$

$$\text{Integrated Response of ERK-PP:} \qquad \sum_i C_i^I = -0.9998.$$

A paradigm of classical biochemistry was that pathways should have a single controlling step. To examine how the control is distributed in EGF–ERK signal transduction, Figure 3.8 displays the control coefficients on amplitude, duration and integrated response. The abscissa describes the reactions $(v_1, \ldots, v_{148})$ and the ordinate indicates the values of the control coefficients. It can been seen that calculated control is far from uniformly

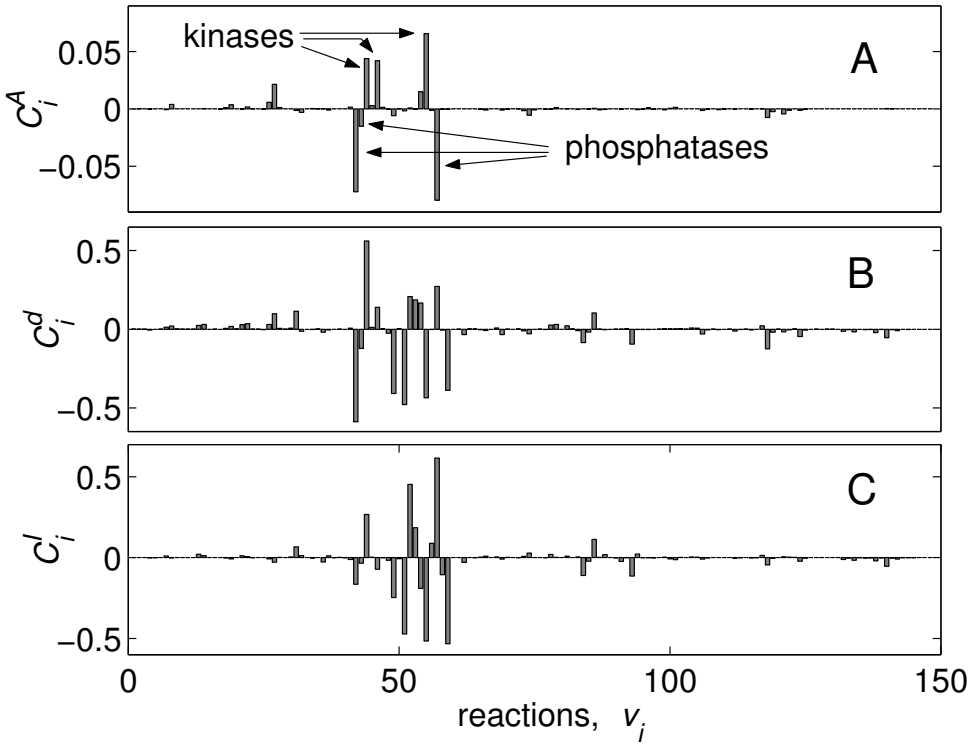


Figure 3.8: Distribution of the three types of control coefficients for ERK-PP calculated for the differentially small perturbations: for the amplitude (A), for the duration (B) and for the integrated response (C). Some control coefficients for reactions involving kinases and phosphatases are indicated in panel A.

distributed over all processes. Most reactions have a control coefficient that is very close to zero and only a small group of reactions show much control. Moreover, the control coefficients for amplitude are about one order of magnitude smaller than the coefficients for duration and integrated response.

For practical purposes a control coefficient is defined to be "substantial" if it is not smaller in absolute value than 10% of the largest control coefficient. That particular threshold value ensures that a moderate number of control coefficients fall into this category and may therefore be considered in more detail. Table 3.1 displays the sub-

stantial positive (marked in red) and the substantial negative (marked in blue) control coefficients for control on the amplitude, the duration, and the integrated response. For instance, with respect to the control on amplitude (small perturbations), only eight out of the 148 processes are found to display substantial control coefficients. These include five reactions with positive control (v_{27}, v_{44}, v_{46}, v_{54}, v_{55}) since their control coefficients are not smaller than 10% of $C^A_{55} = 0.066$. Similarly, only the three reactions (v_{42}, v_{43}, v_{57}) exert substantial negative control since the absolute values of their control coefficients are larger than 10% of the absolute value of $C^A_{57} = -0.08$. In Appendix B.3 the control coefficients of Table 3.1 are depicted superimposed on the biochemical scheme of the signalling network. The trend that becomes apparent is that many of the largest control coefficients are associated with reactions that involve Raf, MEK and ERK. This is irrespective of the system property for which the control distribution is calculated.

A phenomenon which is already discussed for linear signalling cascades in Heinrich et al. [2002] can also be confirmed for the complex model under consideration: Kinases have positive control on different characteristics of the signalling output whereas phosphatases exhibit negative control. An exception is reaction v_{55} (ERK activation by MEK) which exhibits negative control on duration and integrated response. As a reason for this deviation from the control bahaviour of components of linear cascades one might think of the negative feedback which introduces a considerable nonlinearity into the system. However, calculations without this negative feedback mechanism reveal that all control coefficients have the same sign and about the same absolute value (data not shown).

Another aspect which is studied in Heinrich et al. [2002] concerns the question of crosstalk between signalling pathways. Crosstalk is modelled by considering the temporary inhibition of a phosphatase of a pathway by another pathway. There, it can be shown that amplification and short signal duration (spike-like behaviour) can only coexist with this type of crosstalk. Moreover, the further upstream the temporarily inhibited phosphatase is located the more pronounced is this effect. The MAPK model under consideration shows a similar tendency, i.e. the further upstream the phosphatases (see $v_{42/43}$, $v_{49/51}$ and $v_{57/59}$ in Table 3.1), the more control impact on duration and integrated response of ERK they possess.

An further question to address is: Do reactions, which exert a high degree of control on one of the three characteristics of signalling output, tend also to exert a similar influence on the other two characteristics? To address this point, the values of the three types of control coefficients are plotted against each other in Figure 3.9. Figure 3.9C shows that the reactions that have a high control on the duration of the ERK-PP signal

		small pertubations			large pertubations +50%			large pertubations −50%		
v	reactions	A	d	I	A	d	I	A	d	I
26	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[Ras-GDP] ⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]	0.006	0.038	0.070	0.004	0.026	-0.048	0.013	0.070	0.130
27	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP] ⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos] + [Ras-GTP]	0.021	0.125	0.236	0.012	0.089	0.168	0.075	0.221	0.403
31	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP] ⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[Ras-GDP]	-0.001	0.138	0.203	-0.001	0.128	0.192	-0.001	0.151	0.214
42	[Raf*]+[Phosphatase1] ⟶ [Raf*-Phosphatase1]	-0.072	-0.699	-1.130	-0.609	-0.438	-0.927	-0.030	-1.249	-1.965
43	[Raf*-Phosphatase1] ⟶ [Raf]+[Phosphatase1]	-0.015	-0.146	-0.235	-0.013	-0.098	-0.159	-0.017	-0.283	-0.457
44	[MEK] + [Raf*] ⟶ [MEK-Raf*]	0.044	0.640	0.924	0.019	0.570	0.820	0.815	0.667	1.212
46	[MEK-P]+[Raf*] ⟶ [MEK-P-Raf*]	0.042	0.191	0.419	0.019	0.165	0.353	0.696	0.170	0.761
49	[MEK-PP-Phosphatase2] ⟶ [MEK-P] + [Phosphatase2]	-0.006	-0.453	-0.607	-0.006	-0.344	-0.450	-0.006	-0.720	-0.973
51	[MEK-P-Phosphatase2] ⟶ [MEK]+[Phosphatase2]	-0.002	-0.470	-0.440	-0.002	-0.355	-0.341	-0.002	-0.686	-0.609
52	[ERK]+[MEK-PP] ⟶ [ERK-MEK-PP]	0.001	0.202	0.172	0.001	0.161	0.133	0.001	0.286	0.254
53	[ERK-MEK-PP] ⟶ [ERK-P]+[MEK-PP]	0.000	0.184	0.158	0.000	0.119	0.103	0.000	0.402	0.342
54	[ERK-P]+[MEK-PP] ⟶ [ERK-P-MEK-PP]	0.015	0.198	0.327	0.010	0.155	0.266	0.030	0.288	0.442
55	[ERK-P-MEK-PP] ⟶ [ERK-PP]+[MEK-PP]	0.066	-0.404	-0.207	0.038	-0.315	-0.181	0.699	-0.709	0.119
57	[ERK-PP-Phosphatase3] ⟶ [ERK-P]+[Phosphatase3]	-0.080	0.216	-0.091	-0.164	0.184	-0.123	-0.061	0.327	-0.071
59	[ERK-P-Phosphatase3] ⟶ [ERK]+[Phosphatase3]	0.000	-0.389	-0.347	0.000	-0.335	-0.296	-0.001	-0.492	-0.440
74	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP] ⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos] + [Ras-GTPi]	-0.005	-0.043	-0.090	-0.005	-0.037	-0.075	-0.006	-0.052	-0.112
84	[Rafi*]+[Phosphatase1] ⟶ [Rafi*-Phosphatase1]	0.000	-0.086	-0.088	0.000	-0.059	-0.060	-0.002	-0.167	-0.190
86	[MEK] + [Rafi*] ⟶ [MEK-Rafi*]	0.001	0.106	0.110	0.001	0.103	0.108	0.000	0.109	0.112
93	[MEKi-P-Phosphatase2] ⟶ [MEK]+[Phosphatase2]	0.000	-0.089	-0.074	0.000	-0.068	-0.059	0.000	-0.126	-0.099
118	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[Prot]	-0.007	-0.146	-0.216	-0.008	-0.122	-0.182	-0.007	-0.183	-0.270
124	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]+[Prot]	-0.001	-0.054	-0.077	-0.001	-0.047	-0.067	-0.001	-0.065	-0.092
140	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]	0.000	-0.056	-0.059	0.000	-0.054	-0.056	0.000	-0.058	-0.061

Table 3.1: The most substantial positive and the most substantial negative control coefficients on amplitude, duration, and integrated response. The control coefficients calculated with small and large perturbations ($\partial k/k = 10^{-6}$ and $\Delta k/k \pm 0.5$, respectively) are shown. The significantly positive and negative control coefficients are displayed in red and blue respectively.

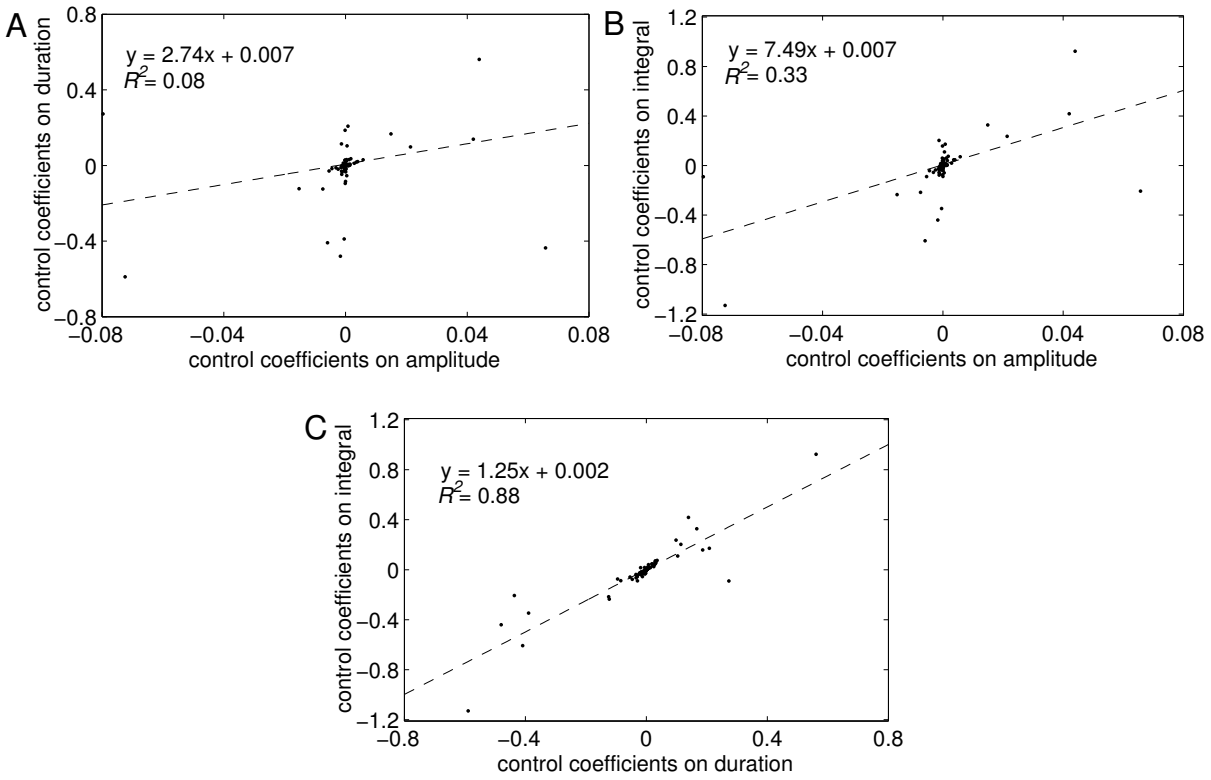


Figure 3.9: Correlations between control coefficients on the three signalling characteristics. A: signal amplitude and signal duration, B: signal amplitude and integrated response, C: signal duration and integrated response. In each case regression lines are depicted (dashed lines), and their linear function and the R-squared value (correlation coefficient) are shown.

do indeed, in addition, tend to exert considerable control on the integrated response of the ERK-PP profile. This is reflected by a relatively high correlation coefficient of $R^2 = 0.88$. In Heinrich et al. [2002] it is shown analytically that in simple linear cascades (without crosstalk effects) a high signal amplitude always coincides with a high signal duration of the output kinase. This implies that both characteristics can not be controlled independently in such cases. In the more complex signalling network under consideration here, these two control coefficients are no longer covariant, as is illustrated by a small correlation coefficient of $R^2 = 0.08$ (see Fig. 3.9A). This indicates that, in contrast to linear cascades, the considered model system possesses the ability to decouple the control of signal amplification and duration.

With respect to drug design the sensitivities of different characteristics of the signal

output might represent an important item of information. To decide if the overall sensitivity of, for example, the amplitude of ERK-PP is significant, this value is compared to the amplitude sensitivities of components which lie further upstream of ERK. To analyse how sensitive or robust the amplitude of a component is on average in terms of parameter perturbations, the measures introduced in Eq. (3.3) and Eq. (3.4) are calculated for amplitudes of various components. Figure 3.10 depicts the distributions of control coefficients for the amplitudes of RasGTP (A), Raf* (B), MEK-P (C), MEK-PP (D), ERK-P (E) and ERK-PP (F). For the six different components, values for σ^A and ρ^A

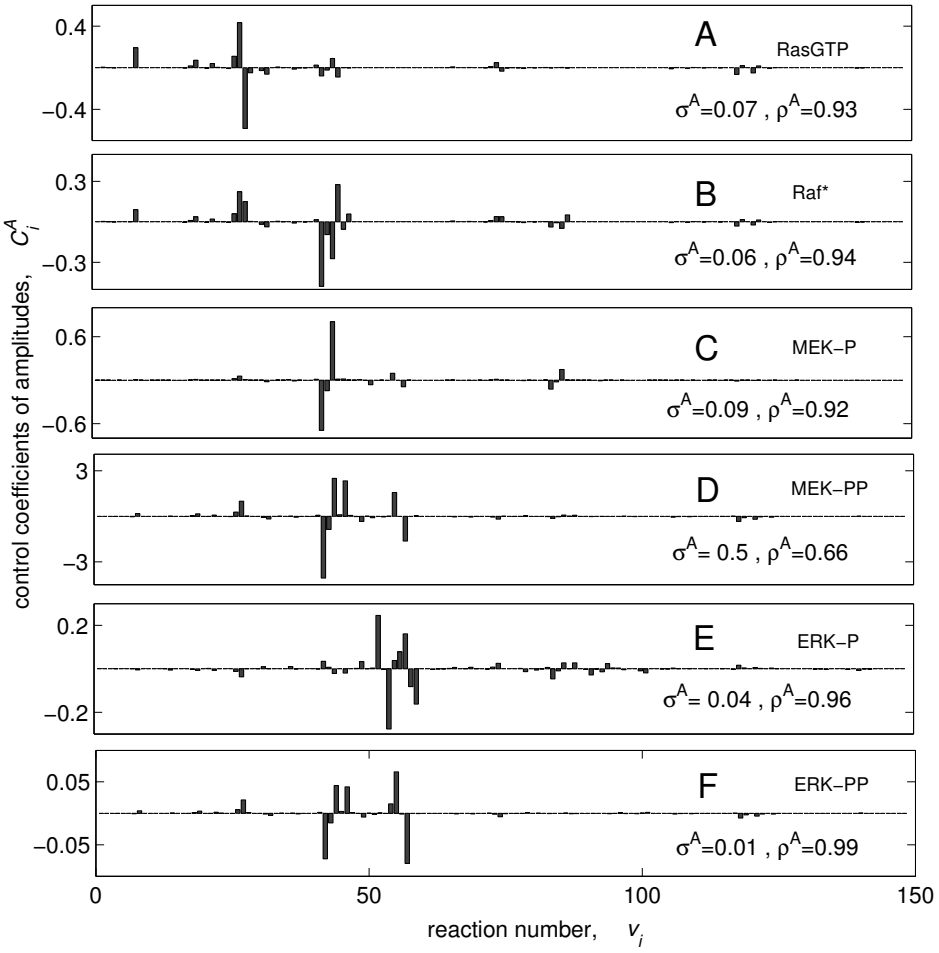


Figure 3.10: Distributions of control coefficients on the amplitude of active Ras (A), Raf* (B), MEK-P (C), MEK-PP (D), ERK-P (E) and ERK-PP (F). Values for overall sensitivities σ^A and overall robustness' ρ^A are shown.

are also indicated. In every case, only few coefficients possess considerably high values. In that respect the distributions of the control coefficients look qualitatively similar.

However, there are big differences in the absolute values. Whereas the signal output ERK-PP is highly robust having very small control coefficients for the amplitude, the active MEK kinase (MEK-PP) has the most sensitive amplitude. The control coefficients of the MEK-PP amplitude vary considerably in the range of $-4.0 < C_i^A < 2.5$. To focus on the processes which yield such a high sensitivity of the amplitude of the active MEK kinase, Figure 3.11 shows a smaller sector of reactions ($40 \leq i \leq 60$) of dual and mono phosphorylated MEK. The control coefficient with the highest value in

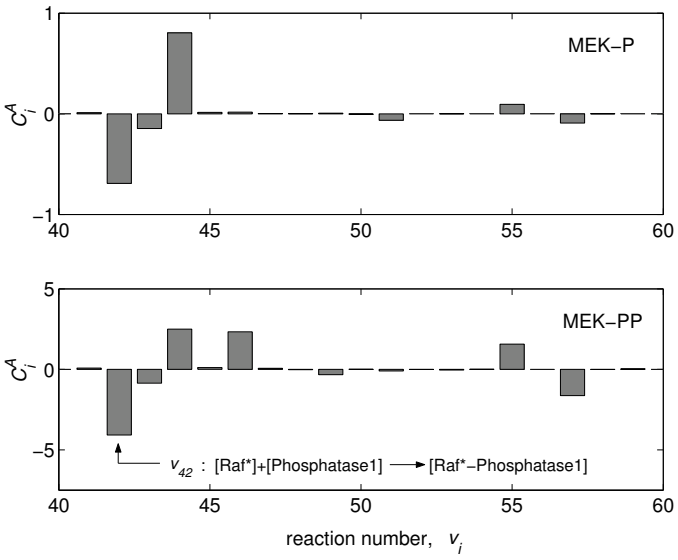


Figure 3.11: Control coefficients for the amplitude of the mono and dual phosphorylated MEK kinase exerted by reactions $v_{40} - v_{60}$. The strongest effect on the amplitude of MEK-PP exhibits reaction v_{42} which is the binding of the first phosphatase to the active Raf component ($C_{42}^A = -4.0$).

absolute terms shows reaction v_{42} describing the binding of active Raf to its specific phosphatase.

The results of this sensitivity analysis carried out with the methods defined in the framework of MCA are as follows: (1) Robust amplitude of the signal output ($\rho^A = 0.99$), i.e. very small control coefficients of the ERK-PP amplitude compared to other kinases and active Ras. (2) The highest overall sensitivity ($\sigma^A = 0.5$) of the cascade "Ras-Raf-MEK-ERK" can be found for the active MEK kinase. (3) Raf* binding to its specific phosphatase exhibits the strongest control on the amplitude of the MEK activation profile ($C_{42}^A = -4.0$).

The control coefficients are calculated by applying small perturbations to the rate

constants. To investigate how strongly the size of the perturbation affects the distribution of control, the relationships between the large and small perturbations are plotted in Figure 3.12. The filled circles indicate the relationship of perturbations with $\Delta k/k = 0.5$ and $\partial k/k = 10^{-6}$, the blank circles accordingly the interrelation between perturbations with $\Delta k/k = -0.5$ and $\partial k/k = 10^{-6}$. Were the control coefficients invariant with respect to the size of the perturbation (which would hold for linear systems), these plots would show a linear relationship with a slope of 1. Here, the slope of the relation is not

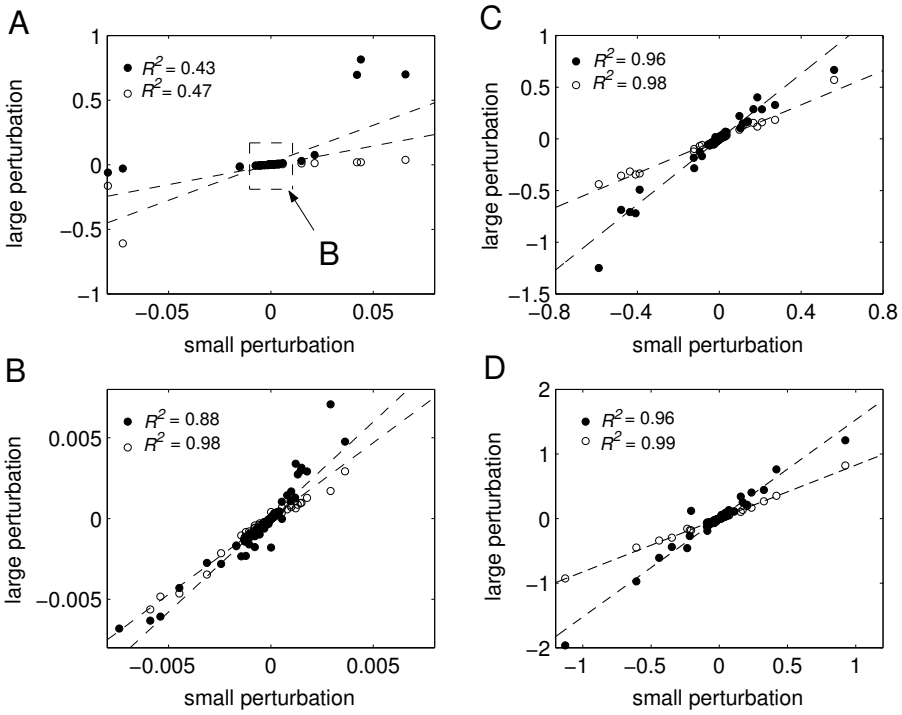


Figure 3.12: Relationship between large and differentially small perturbations. For each process, the value of the control coefficient calculated by large perturbations (blank circles for $\Delta k/k = -0.5$, filled circles for $\Delta k/k = +0.5$) is plotted against the value of the control coefficient calculated by a small perturbation ($\partial k/k = 10^{-6}$): amplitude (A and B), duration (C), integrated response (D). Figure B is a magnification of the dashed box in Figure A. The correlation between both ways of calculating the control coefficients is determined by linear regression (dashed lines), which yields in each case a R-squared value.

1, but the points approximate a straight line and the correlation between the control coefficients calculated by small and those calculated by large perturbations is close to

1, except for the coefficients for the amplitude (Figure 3.12A). However, this is caused by a few outliers with high values. If these are not taken into account, the correlation for amplitude is also fairly reasonable (Figure 3.12B). This means that the size of the control coefficients is indeed affected by the size of the perturbation, but, importantly, the order in which the processes are ranked on the basis of their control coefficients does not change.

The control considered so far is that by molecular and biochemical processes. This does not necessarily indicate the importance of proteins. The question of how changes in the concentration of the proteins (caused by e.g. changes in gene expression) influence signalling through the network is also addressed. For this, the fractional change in the three signalling characteristics caused by a fractional change in the total concentration of any protein P_{total} is calculated. This measure corresponds to the response coefficient of Metabolic Control Analysis:

$$R^{y}_{P_{\text{total}}} = \frac{\partial \ln y}{\partial \ln P_{\text{total}}} \qquad \text{for y } \in \{A, d, I\}.$$

A further application to another specific signal transduction pathway is given in Lee et al. [2003]. Table 3.2 shows the coefficients quantifying the responses of the three characteristics of signalling to ERK to changes in the concentrations of the proteins in the network.

Again a "substantial" response coefficient is defined as one for which the absolute value exceeds 10% of the highest response coefficient. Signalling duration and the integrated output are affected most by changes in the concentrations of Ras, MEK and ERK (positive effect) and the phosphatases 1, 2 and 3 (negative effect). When small changes in protein concentrations are applied, the amplitude is, surprisingly, sensitive solely to the concentration of ERK and its phosphatase. When large changes in protein concentrations are considered, the amplitude also becomes sensitive to an increase in the concentration of Phosphatase1 (negative effect) as well as to a decrease in the concentrations of Ras and MEK (positive effect). Changes in the concentrations of EGF, the adaptor proteins and Raf do not seriously affect the output of MAPK signalling.

Hitherto, specific signalling pathways have been considered. In section 3.4 a broader perspective is adopted by systematically investigating cascades of varying lengths and phosphorylation mechanisms due to a hypothetical objective function (overall sensitivity of the amplitude of the signal output).

	small pertubations			+50% pertubations			−50% pertubations		
protein	A	d	I	A	d	I	A	d	I
EGF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EGFR	0.01	0.17	0.28	0.00	0.11	0.20	0.02	0.27	0.40
Prot	-0.02	-0.32	-0.49	-0.02	-0.25	-0.37	-0.01	-0.51	-0.78
shc	0.00	-0.01	0.00	0.00	-0.03	-0.03	0.00	0.02	0.03
Grb2	0.00	-0.01	-0.01	0.00	-0.01	-0.01	0.00	-0.01	-0.01
Sos	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.02	0.03
RasGDP	0.08	0.84	1.30	0.03	0.73	1.12	1.95	1.07	1.97
GAP	0.02	0.21	0.32	0.01	0.10	0.18	0.06	0.40	0.56
grb2sos	0.00	-0.02	0.00	0.00	-0.02	-0.02	0.01	0.01	0.04
Raf	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.03
Phosphatase1	-0.09	-0.95	-1.47	-1.25	-0.56	-1.40	-0.03	-1.64	-2.57
MEK	0.05	0.90	1.24	0.02	0.88	1.18	1.14	0.85	1.49
Phosphatase2	-0.01	-1.03	-1.12	-0.01	-0.70	-0.76	-0.01	-1.94	-2.08
ERK	1.91	0.70	2.74	1.91	0.53	3.15	1.85	1.18	1.94
Phosphatase3	-0.94	-0.49	-1.67	-1.02	-0.53	-1.35	-0.93	-0.47	-2.10

Table 3.2: Response coefficients $R^{A,d,I}_{P_{\text{total}}}$. The relative importance of the concentration of each protein or protein complex is calculated with small (actual response coefficient, $\partial P_{\text{total}}/P_{\text{total}} = 10^{-6}$) and large ($\Delta P_{\text{total}}/P_{\text{total}} = \pm 0.5$) perturbations. The substantially positive and negative response coefficients are displayed in red and blue, respectively.

3.4 Sensitivity and optimal length of signalling cascades

In the present section the concept of an overall sensitivity defined within the framework of MCA (Eq. 3.3) is applied systematically on kinase cascades. A different kind of sensitivity analysis of kinase cascades has already been performed for *steady states* (Goldbeter and Koshland [1981]; Goldbeter and Koshland [1982]; Huang and Ferrell [1996]). Particularly, the impact of multistep effects (Goldbeter and Koshland [1982]) and distributive dual phosphorylation (Huang and Ferrell [1996]; Ferrell and Bhatt [1997]) on an ultrasensitive response of the signal output have been investigated in these studies. One result states that the longer the kinase cascades are, the stronger they exhibit zero-order ultra-sensitivity.

The aim of the approach presented here concerns about a possible optimal length n of a cascade in which optimality is quantified as a maximum overall sensitivity measured in terms of control coefficients for the signal amplitude of the last kinase. The basic

question is, if the average overall sensitivity σ^A of the amplitude of the signal output is an appropriate objective function for signalling cascades. Concerning the dependence of σ^A on the cascade length, a special focus lies moreover on differences between mono phosphorylation cascades and more realistic designs, i.e. dual phosphorylation cascades. Such an approach implicitly associates a high overall sensitivity of the signal output with a high controllability thereof.

3.4.1 Model description

Linear signal transduction pathways each consisting of a receptor with concentration $R(t)$ ($\equiv X_0(t)$) and up to n kinases and phosphatases are considered. Similarly to chapter 2, mass action kinetics are used and phosphatases are not modeled explicitly within the different pathway models. The calculations are performed for monocyclic cascades schematically depicted in Figure 3.13A and for dual phosphorylation cycles depicted in Figure 3.13B. However, also for dual phosphorylation cascades the very first cycle is always assumed to be a mono phosphorylation step. The reason for this is that "real" MAPK cascades indeed show this particular property of a single phosphorylation of the first kinase (MAPKKK) to be active. Furthermore, this assumption enables a better comparison between both types of cascades when enlarging them from a one step pathway, having the same sensitivity properties, to longer pathways ($2 \leq n \leq 5$).

The pathways are modeled with ODEs similarly to the rate equations introduced in section 2.1.1. Provided that the kinases obey conservation relations within the phosphorylation/dephosphorylation cycles one gets the following equations for kinases in monocycle cascades:

$$\frac{dX_i^*}{dt} = \alpha_i X_{i-1}^* (1 - \frac{X_i^*}{C_i}) - \beta_i X_i^* \qquad \text{for } i = 1 \ldots n. \tag{3.9}$$

The model quantities are as follows:

- X_i^*, X_i: concentrations of activated (mono phosphorylated) and inactivated kinases (unphosphorylated), respectively.
- $C_i = X_i + X_i^*$: total concentration of kinase i.
- α_i, β_i: rate constants for kinases and phosphatases, respectively.

The dual phosphorylation steps are assumed to occur by a distributive mechanism. This means that the dual phosphorylation of all kinases (except the first kinase) proceeds through a two-collision mechanism rather than a single-collision processive mechanism

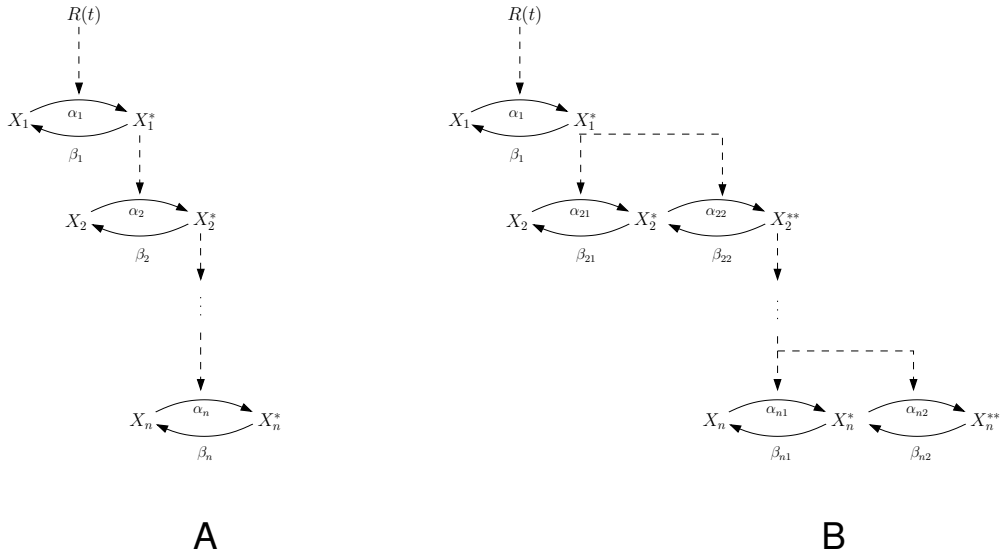


Figure 3.13: Linear signalling pathways of different length n with mono phosphorylation (A) and dual phosphorylation (B) cycles. System variables for pathways of type A are X_i and X_i^* indicating the concentrations of activated (mono phosphorylated) and inactivated (unphosphorylated) kinases, respectively. System variables for pathways of type B are X_i, X_i^* and X_i^{**}, representing the concentrations of activated (dual phosphorylated) and inactivated kinases (mono phosphorylated, unphosphorylated), respectively. In either case the first activation step is modeled as a mono phosphorylation step. A transient stimulation of the receptor is described by $R(t) = 1 \cdot \exp(-\lambda t)$, with $\lambda = 1$.

(Ferrell and Bhatt [1997]). The arising scheme of such a cascade is depicted in 3.13B and the equation system is given as

$$\begin{aligned}
\frac{dX_1^*}{dt} &= v_1 - v_{-1} = \alpha_1 X_0 (1 - \frac{X_1^*}{C_1}) - \beta_1 X_1^*, \\
\frac{dX_2^*}{dt} &= v_{21} - v_{22} - v_{-21} + v_{-22} \\
&= \alpha_{21} X_1^* (1 - \frac{X_2^* + X_2^{**}}{C_2}) - \alpha_{22} X_2^* X_1^* - \beta_{21} X_2^* + \beta_{22} X_2^{**}, \\
\frac{dX_2^{**}}{dt} &= v_{22} - v_{-22} = \alpha_{22} X_2^* X_1^* - \beta_{22} X_2^{**}, \\
\frac{dX_i^*}{dt} &= v_{i1} - v_{i2} - v_{-i1} + v_{-i2} \\
&= \alpha_{i1} X_{i-1}^{**} (1 - \frac{X_i^* + X_i^{**}}{C_i}) - \alpha_{i2} X_i^* X_{i-1}^{**} - \beta_{i1} X_i^* + \beta_{i2} X_i^{**}, \\
\frac{dX_i^{**}}{dt} &= v_{i2} - v_{-i2} = \alpha_{i2} X_i^* X_{i-1}^{**} - \beta_{i2} X_i^{**} \qquad \text{for } i = 3 \ldots n,
\end{aligned} \tag{3.10}$$

where the model quantities are as follows:

- X_i^{**}, X_i^*, X_i: concentrations of activated (dual phosphorylated) and inactivated kinases (mono phosphorylated, unphosphorylated), respectively.

- $C_1 = X_1 + X_1^*$, $C_i = X_i + X_i^* + X_i^{**}$: total concentrations of the first kinase ($i = 1$) and all other kinases ($i \geq 2$), respectively.

- α_{i1}, α_{i2}, β_{i1}, β_{i2}: rate constants of kinases and phosphatases.

The impact of the structural design (mono/dual phosphorylation and cascade length n) on the sensitivity of the signal amplitude (σ^A) is considered, whereas appropriately the amplitude is defined as the maximum activation level of X_n^* in the case of monocylces and X_n^{**} in the case of dual phosphorylation steps. To simplify this structural analysis of signalling cascades, the total concentrations of all kinases and the rate constants are assumed in either case to be equal ($C_i = C = 100$ and $\alpha_{i(j)} = \alpha$, $\beta_{i(j)} = \beta$). Therefore, one can rescale the time in the equation systems dividing all equations by α which leaves only one kinetic parameter in the systems ($\beta/\alpha \equiv \beta$). The receptor $R(t) = X_0(t)$ is assumed to be initially activated at time $t = 0$ and then decays exponentially: $X_0(t) = 1 \cdot \exp(-\lambda t)$ with $\lambda = 1$.

3.4.2 Results

The number of reaction rates and therefore the number of control coefficients of a given cascade depend on the length n ($1 < n < 5$) and on the phosphorylation mechanism. For each cascade design, control coefficients are calculated numerically ($\partial v/v = 10^{-6}$). Figure 3.14 depicts how the overall sensitivity of the amplitude of the output kinase depends on the cascade length n, the scaled phosphatase rate constant β and the phosphorylation mechanism.

The left panels (Fig. 3.14A) refer to mono phosphorylation cascades and the right panels (Fig. 3.14B) depict the situation for dual phosphorylation pathways. As examples, on top of each row the activation profiles of the last activated kinases of five-step cascades, i.e. X_5^* (left) and X_5^{**} (right) are depicted for three different β-values. The bars in the histograms show the overall sensitivity of the last kinase (X_n^* or X_n^{**}) in a pathway with length n. Two gray shadings indicate different amplification characteristics of a given pathway. The amplification/dampening criteria are the same as in section 3.1: If the activation profile of the last kinase $X_n^{*(*)}$ of a given pathway has a maximum value higher than the initial value of the receptor which is set to $X_0(t = 0) \equiv 1$, the pathway

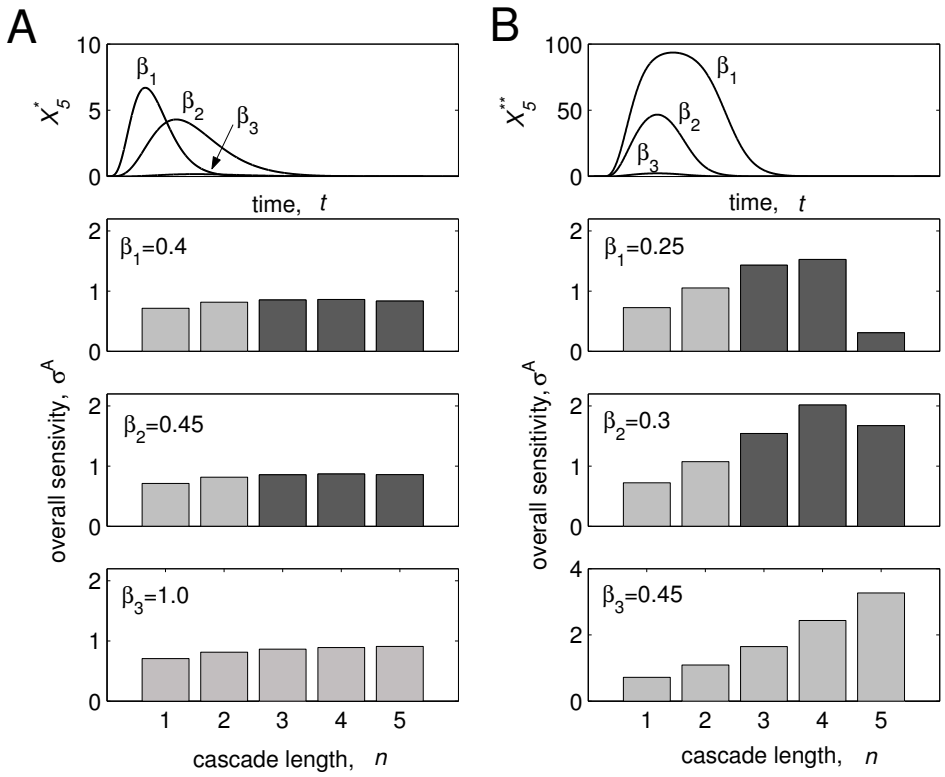


Figure 3.14: Dependence of overall amplitude sensitivity σ^A of mono phosphorylation cascades (A) and dual phosphorylation cascades (B) on length n and β. In each case the first activation step is a mono phosphorylation cycle. The upper two panels show as examples the activation profiles of the last activated kinases of five-step cascades, i.e. X_5^* (left) and X_5^{**} (right), in each case for three different β-values. Other parameters are: $C_i = C = 100$ and $\lambda = 1$. Dark gray shaded bars indicate that a given cascade amplifies signals: $\max\{X_n^{*(*)}\} \geq X_0(t = 0)$.

is assumed to be amplifying. Otherwise the pathway is dampening. Dark gray color indicates an amplifying and light gray a dampening cascade with length n.

Overall sensitivities σ^A are calculated for three different β-values for each of the two phosphorylation mechanisms. For both mechanisms, β_1 and β_2 are chosen such that the cascade ultimately becomes amplifying the longer it gets. If the phosphatase activity is too high (e.g. β_3), the cascades get more and more dampening the longer they become.

It can be seen that for mono phosphorylation cascades (panels in Figure 3.14A) the sensitivity of the amplitude depends only moderately on the cascade length n. This is irrespective of the β value. In contrast to that, dual phosphorylation cascades show a

qualitatively different behaviour. Here, the sensitivity depends much stronger on the cascade length. A maximum sensitivity for moderate cascade length exists ($n^{\max} = 4$ in the depicted examples) for β-values which enable the cascade to amplify signals, but it is most significant for $\beta = 0.25$. In fact, if phosphatase activities are too high (e.g. $\beta = 0.45$), the sensitivity increases monotonically with increasing length n (lower panel of Figure 3.14B). However, the signalling cascades for such high β-values become more and more dampening and are therefore not considered as values proper for functioning signalling cascades.

The existence of a cascade length $n^{\max}$ (in an appropriate range of β-values) with a *pronounced* maximum sensitivity is related to the property of dual phosphorylation cascades to strongly amplify signals. In contrast to the signal output of mono phosphorylation cascades, for small n the amplitude of the last kinase X_n^{**} gets *rapidly* more amplified provided that the β-values allow for amplification (see upper panel of Figure 3.14B). Depending on the values for β and the total concentration of kinases, a cascade length exists (in the examples of Figure 3.14B one derives $n = n^{\max} + 1 = 5$) for which the signal is highly amplified and almost all molecules of the last kinase are activated ($X_n^{**} \approx C_n$). This makes the cascade on average less sensitive if it is enlarged further, because the control coefficients on the amplitude of the signal output are getting smaller in this case. Together with an increase of σ^A-values for n-values until $n^{\max}$, one derives a distinct maximum for the overall sensitivity of the amplitude.

3.5 Discussion

In this chapter, principles behind the multifarious control of signal transduction are investigated. The demonstrated principles confirm and elaborate on previous predictions that have been made on the basis of a theoretical analysis of a simple model of signal transduction (Heinrich et al. [2002]). For the approach used in the present chapter, the application of Metabolic Control Analysis to transient activation profiles of signalling pathways plays a central role .

In a first step, the combined analysis of a simple kinetic model of a signalling pathway with mathematics and with quantitative experimentation is described. Using such a mathematical model, the effects of kinase and phosphatase inhibitors can be simulated and compared with the results of inhibitor experiments. An important question of this approach is how the control of kinases and phosphatases, exerted on the activation profile of the signal output, differs. Of particular interest are four features of the activity of a signalling cascade: the amplitude, the final steady state level, the duration of signalling

and the integrated response.

The experimental results and the simulations correspond well, particularly when high kinase activations that lead to saturation effects, are considered. The findings illustrate that a kinase (MEK) inhibitor affects the amplitude of signalling through MAPK, while leaving duration rather unaffected. A (protein tyrosine) phosphatase inhibitor influences both the duration of the MAPK activation profile and its final level in the steady state. The effect of the phosphatase inhibitor to the ERK-PP amplitude is considerably lower. In the case of saturation, this property is more pronounced which can be attributed to the fact that EGF stimulation causes virtually all ERK molecules to be dual phosphorylated when the maximum is reached. Phosphatase inhibition can therefore not further elevate the amplitude in the peak. The findings that phosphatases and kinases have different control effects on amplitude and duration of the signal output, are closely related to the summation laws of these two dynamic properties. Since the control coefficients of the signal duration sum up to -1, components with negative control (e.g. phosphatases) cause a higher effect thereon.

In a second step, control analysis is applied to a detailed kinetic model of EGF-induced MAPK signalling (Schoeberl et al. [2002]). Some components of the MAPK network are well-known oncogenes, such as Ras and Raf, causing the pathway to be constitutively active in tumour cells (Hoshino et al. [1999]), and have therefore been taken to be promising targets for rationalised therapies against cancer (Sebolt-Leopold [2000]). For each of the 148 processes in the network, it is calculated to what extent the process controls the shape of the ERK-PP profile in terms of amplitude, duration and integrated response. Most reactions do not control these characteristics at all or only to a very small extent. This striking result might help identifying candidate reactions suitable as targets for cancer therapy.

The reactions with the highest control coefficients with respect to all characteristics of the ERK-PP profile involve binding of Raf, MEK and ERK to each other and to their respective phosphatases. Furthermore, it can be observed that upstream dephosphorylation processes exhibit higher control on the duration and integrated response of ERK-PP than downstream dephosphorylation processes. This phenomenon might be of interest in the context of possible crosstalk effects, specifically the temporary inhibition of a phosphatase, resulting from a pathway running in parallel to the MAP kinase cascade. Heinrich et al. [2002] shows analytically that amplification and short signal duration (spike-like behaviour) can only coexist with this type of crosstalk. Additionally, under certain conditions (strongly activated pathways), the more upstream a slow phosphatase is located, the stronger the control effect on signal duration. It might be of interest for future projects to figure out appropriate target phosphatases for inhibition,

either for a deeper understanding of crosstalk principles or for drug design.

In contrast to the high robustness of the ERK-PP amplitude, reflected by very small control coefficients, the MEK kinase is highly sensitive. It remains to be elucidated, if there is a connexion between this property and the well-known fact that MAPK kinases are comparable specific activators for their target kinases (Cowan and Storey [2003]).

Last, the systematic analysis of signalling pathways with respect to the sensitivity of the signal output represents a first step towards extending previous investigations (Goldbeter and Koshland [1981]; Goldbeter and Koshland [1982]; Huang and Ferrell [1996]; Ferrell and Bhatt [1997]) in two directions:

- Instead of considering steady states, the more interesting operation regime of signal transduction represented by transient activation profiles is analysed.

- As an objective function, the overall sensitivity σ^A (in contrast to zero-order ultrasensitivity) is considered to assess why longer, dual phosphorylation cascades are advantageous compared to mono phosphorylation cascades.

So far, the longer the cascades are, the more they have been considered to be advantageous regarding their ultrasensitive behaviour. Therefore, another basic question is, whether there is an optimal length with respect to the criteria chosen in the present study. It is still a matter of debate why there are exactly three kinases in sequence in a MAP kinase cascade. The strategy followed here is open for at least two particular further investigations:

- Other objective functions like sensitivity/robustness of signal duration, signalling time, integrated response, etc. might be considered. Since response coefficients often play an important role in signalling, objective functions may also be considered based on them.

- As the total concentrations of participating kinases C_i play an intriguing role for the overall sensitivity of signalling pathways, it might be of particular interest to analyse the impact of the *ratios* of these total kinase concentrations. This may be done by investigating, whether optimal ratios of C_i-values exist. Such values might then be compared to kinase concentrations found in the literature or databases.

The latter aspect has already been investigated for two-component signal transduction systems which are frequently used by bacteria to sense environmental conditions. These systems consist of a sensor and a response kinase. For the KdpD/KdpE system of E. coli, an optimal signal transduction is guaranteed, if the concentration of the second kinase (response kinase) is about 30 times higher than the first (sensor) kinase (Saez-Rodriguez et al. [2004]).

Chapter 4

Conclusions

In contrast to chemical reaction systems of inanimate nature, biochemical reaction systems as metabolic networks, signal transduction or gene regulatory networks are the outcome of evolution. This means that their design is not arbitrary, but such that the networks may fulfill their specific functions for living cells in an optimal way (Heinrich et al. [1991]; Stephani et al. [1999]). It is therefore of interest to analyse which kind of structures bring about certain advantageous dynamic features. Hitherto, studies of that kind have been performed only for metabolic networks (Stephani et al. [1999]; Ebenhöh and Heinrich [2003]) resulting in conclusions concerning an optimal ordering of reactions in pathways of energy metabolism as well as concerning their robustness against elimination or exchange of reactions.

The present work extends this kind of studies to signalling networks. Two approaches are presented following a similar perspective. They are directed to examine the structural design and control properties of signal transduction systems using various criteria to evaluate certain design and control principles.

In chapter 2, large sets of alternative structural designs for signalling networks of small size ($n \leq 7$) have been analysed analytically. The objective is to find out what kind of interrelations exist between specific dynamic properties such as signal amplification or dynamic stability and design principles. One finds several hundred kinases in living cells. Therefore, the aim of the strategy based on small networks is to predict an optimal design – in terms of network connectivity and number of cycles – of such large signalling networks. It can be demonstrated that a large kinase network has to exhibit a low network connectivity and a low number of cycles in order to avoid autoactivation due to spurious receptor or kinase stimulations. Networks retrieved from the TRANSPATH database indeed display these features.

Investigations of the structural robustness are carried out for networks with four kinases. Two values of phosphatase activities (β/α), for which an ensemble of amplifying networks is most robust, can be determined. It would be interesting for future projects to extend these calculations to larger networks. This would yield β/α-values that might then be compared to real ratios of phosphatase and kinase activities.

The TRANSPATH network exhibits many characteristics which differ significantly from according features in random networks of the same size. This includes the number of pathways and the composition of pathway lengths which can be encountered. Furthermore, it can be shown that the type of crosstalk between pathways is different in the TRANSPATH network compared to random networks, where one finds a normal distribution of frequencies of crosstalk strengths. It seems as if in a real signalling network, crosstalk strengths among pathways are not chosen arbitrarily. Concerning the use of random networks as null hypothesis, other methods for creating these networks might be considered in future projects as well.

Regarding the dynamic characteristics of the design of the TRANSPATH network, there is a tendency that kinase amplitudes are less homogeneously distributed as expected from random simulations. Only few exhibit a considerable signal amplification which can be associated with kinases involved in protein translation.

The TRANSPATH network is also evaluated regarding its capacity to avoid feedback cycles that may occur during an evolutionary process. For this purpose, random networks with the same node degree are generated where small cycles have been avoided. Random networks of such type are not interpreted as null hypothesis, but as structural designs which prevent successfully the emergence of feedback cycles at all by avoiding small cycles with certain mechanisms. Scaffold proteins might be the basis for such a mechanism. The design of other biological networks may be analysed with respect to their potential to avoid feedback cycles. For example, in gene regulation systems, positive feedback cycles seem to play a more pronounced role and therefore emerge by mutations more easily.

The presented strategy for analysing signalling networks is a new conceptual approach. It can be demonstrated that design principles of the TRANSPATH network are indeed distinguishable from those of random networks. The plan is to extend the theory and its applications to networks including multiple phosphorylation of kinases and to consider their interaction with scaffolds or transcription factors. More recent information from TRANSPATH and other databases (e.g. Gough [2002]) shall be taken into account to update the present network structure.

The approach presented in chapter 3 contributes to the comprehension of control

principles within signalling networks. For this purpose, the framework of Metabolic Control Theory is used.

The outcome of inhibitor experiments for kinases and phosphatases are compared with a simple EGF signalling model. Here, it can be demonstrated that phosphatases have a strong effect on the signal duration of ERK and kinases exert high control on the amplitude.

Furthermore, control coefficients of the transient activation profiles of kinases and the G-protein Ras are calculated on the basis of a large EGF model. Regarding the distribution of coefficients, only few processes exert significant control on the signal characteristics (amplitude, duration, integrated response) of these components. Whereas the signal output (ERK-PP) possesses a quite robust amplitude, which is expressed in very small control coefficients, the MEK kinase is much more sensitive to parameter perturbations. It might be of interest to analyse, whether the latter result (highly sensitive MEK) correlates with the finding that MAPK kinases are more specific towards their target substrates than upstream and downstream kinases. Generally, phosphatases exert higher control on signal duration of the signal output than kinases. The impact on duration is more pronounced, the more upstream the phosphatase is. Many control principles which are revealed analytically for linear cascades in Heinrich et al. [2002] can be confirmed for the complex model on hand.

In section 3.4, the question is raised, if appropriate objective functions can be conceptualized for signalling pathways equivalent to objective functions in metabolic systems (e.g. flux). It is systematically tested, whether the overall sensitivity of certain characteristics of the signal output is an appropriate objective function. In this respect, there are qualitative differences between monocycle cascades and dual phosphorylation cascades. The latter exhibit indeed – under certain conditions – extreme values of the sensitivity for moderate cascade lengths. These studies are still preliminary and should be extended to more complex structural designs as well as other possible objective functions.

The insight gained concerning the fundamental design and control principles of cellular signalling may serve as a foundation for future studies in a variety of research areas in the field of signal transduction.

Appendix A

Additional topics to chapter 2

A.1 Explicit solutions for signalling characteristics in low activated networks

Under conditions of weakly activation ($X_i^*/C_i \rightarrow 0$, for each $i = 1 \ldots n$), Eq. (2.3) simplifies to

$$\frac{dX_i^*}{dt} = \rho_i R(t) + \sum_{k \neq i}^{n} \alpha_{ik} X_k^* - \beta_i X_i^*. \tag{A.1}$$

For $\alpha_{ik} = \alpha$ and $\beta_i = \beta$ this equation can be written in a more compact form as follows

$$\dot{\vec{X}}^* = \vec{R}(t) + \mathbf{A}^T \cdot \vec{X}^* - \frac{\beta}{\alpha}\vec{X}^*, \tag{A.2}$$

where the vector $\vec{R}(t)$ has only entries at positions i for those kinases i which do not possess an activation input from other kinases. This specific receptor attachment guarantees all kinases to be activated after receptor stimulation. $\mathbf{A}^T$ is the transposed of the unweighted adjacency matrix with entries $a_{ik} = 1$ if kinase i activates kinase k. The signalling network is in an off-state at the beginning ($t = 0$) and the receptor ($R_i(t) = R_i(0)e^{-\lambda t}$) declines exponentially for $t \rightarrow \infty$. Integration of Eq. (A.2) yields a recurrent formula for the integrated response:

$$\begin{aligned} \int_0^\infty \dot{\vec{X}}^* dt &= \frac{\vec{R}(0)}{\lambda} + \mathbf{A}^T \cdot \int_0^\infty \vec{X}^* dt - \frac{\beta}{\alpha}\int_0^\infty \vec{X}^* dt \\ 0 &= \frac{\vec{R}(0)}{\lambda} + \mathbf{A'}^T \vec{I}, \end{aligned} \tag{A.3}$$

where $\mathbf{A'}^T$ now describes the transposed adjacency matrix of the underlying directed graph with additional β/α-values as entries on the diagonal. For the integrated response

$\vec{I}$ one calculates the following expression:

$$\vec{I} = -(\mathbf{A}'^T)^{-1}\frac{\vec{R}(0)}{\lambda}. \tag{A.4}$$

Multiplication of Eq. (A.2) with time t and t^2 and subsequent integration yield the expressions

$$\begin{aligned}
\int_0^\infty t\dot{\vec{X}}^* dt &= \frac{\vec{R}(0)}{\lambda^2} + \mathbf{A}'^T \cdot \underbrace{\int_0^\infty t\vec{X}^* dt}_{\equiv \vec{T}} \\
-\int_0^\infty \vec{X}^* dt = -\vec{I} &= \frac{\vec{R}(0)}{\lambda^2} + \mathbf{A}'^T \cdot \vec{T} \\
\Rightarrow \vec{T} &= -\left(\frac{\vec{R}(0)}{\lambda^2} + \vec{I}\right) \cdot (\mathbf{A}'^T)^{-1}
\end{aligned} \tag{A.5}$$

and

$$\begin{aligned}
\int_0^\infty t^2\dot{\vec{X}}^* dt &= \frac{\vec{R}(0)}{\lambda^3} + \mathbf{A}'^T \cdot \underbrace{\int_0^\infty t^2\vec{X}^* dt}_{\equiv \vec{Q}} \\
-2\cdot\int_0^\infty t\dot{\vec{X}}^* dt = -2\cdot\vec{T} &= \frac{\vec{R}(0)}{\lambda^3} + \mathbf{A}'^T \cdot \vec{Q} \\
\Rightarrow \vec{Q} &= \left(-2\cdot\vec{T} - \frac{\vec{R}(0)}{\lambda^2}\right) \cdot (\mathbf{A}'^T)^{-1},
\end{aligned} \tag{A.6}$$

respectively. On the left side integrations by parts have been applied.

For a given network design (via the adjacency matrix $\mathbf{A}$) and a given kinetic parameter (β/α) one can calculate $\vec{I}$, $\vec{T}$ and $\vec{Q}$ with Eqs. (A.4–A.6). Signalling amplitude, signalling time and signalling duration of activation profiles of all kinases in the network may then be calculated without using the explicit solutions X_i^* of the differential equation system A.1:

$$\tau_i = \frac{T_i}{I_i} \quad , \quad \vartheta_i = \sqrt{\frac{Q_i}{I_i} - \tau_i^2} \quad , \quad S_i = \frac{I_i}{2\vartheta_i}. \tag{A.7}$$

The numbers of all possible network designs of a given size n are listed in the next section A.2.

A.2 Numbers of network designs

We are interested in network designs without isolated nodes (kinases) and where the edges (representing activations) are directed. If in such networks it is possible to reach any node starting from any other node by traversing edges in some direction (i.e., not necessarily in the direction they point) they are called weakly connected graphs. Therefore, the nodes in a weakly connected graph must all have either outdegree or indegree of at least 1. The numbers of weakly connected graphs with n nodes and e edges can be calculated according to Pòlya's enumeration theorem (Harary [1994]). They are listed in the following Table A.1.

e	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
1	1	0	0	0	0	0
2	1	3	0	0	0	0
3		4	8	0	0	0
4		4	22	27	0	0
5		1	37	108	91	0
6		1	47	326	582	350
7			38	667	2,432	3,024
8			27	1,127	7,694	17,314
9			13	1,477	19,646	74,676
10			5	1,665	42,148	266,364
11			1	1,489	77,305	808,620
12			1	1,154	122,953	2,144,407
13				707	170,315	5,020,543
14				379	206,982	10,490,367
15				154	220,768	19,667,480
16				61	207,301	33,269,008
17				16	171,008	50,927,106
18				5	124,110	70,762,906
19				1	78,813	89,382,317
20				1	43,862	102,797,259

21					21,209	107,693,823
22					8,951	102,832,176
23					3,242	89,457,910
24					1,043	70,886,056
25					288	51,098,059
26					76	33,476,837
27					17	19,889,497
28					5	10,699,652
29					1	5,194,349
30					1	2,271,935
31						890,959
32						313,312
33						98,161
34						27,649
35						6,940
36						1,637
37						346
38						79
39						17
40						5
41						1
42						1
sum	2	13	199	9,364	1,530,843	880,471,142

Table A.1: Number of structural designs with different numbers of kinases n and interactions e.

A.3 Existence of autoactivated states for networks containing cycles

For $R = 0$ the time independent solutions of Eq. (2.3) can be attained by solving the equation system

$$\dot{\vec{X}}^* = 0 = \mathbf{A}^T \cdot \vec{X}^* \cdot \mathbf{D} - \frac{\beta}{\alpha}\vec{X}^*, \quad \text{with} \quad 0 \leq X_i^* < C_i \qquad \forall i \in \{1, 2, \ldots, n\}. \tag{A.8}$$

The matrix $\mathbf{D} = \mathrm{diag}(C_1 - X_1^*, \ldots, C_n - X_n^*)$ is a diagonal matrix and $\mathbf{A}^T$ is the transposed of the adjacency matrix.

The nodes of any graph can be sorted such that the topology of the graph is not affected. Applying this topological sorting to a directed graph without cycles one can arrange the nodes such that the resulting isomorphic graph has an adjacency matrix $\mathbf{A}_{\mathrm{ord}}$ with entries only in the upper-triangle part. This is due to the fact that the lower triangle represents "backward" edges from higher numbered to lower numbered nodes and those edges do not exist in a directed acyclic graph (DAG). Considering a network without feedback cycles, Eq. (A.8) can be rewritten with an ordered adjacency matrix:

$$0 = \mathbf{A}_{\mathrm{ord}}^T \cdot \vec{X}^* \cdot \mathbf{D} - \frac{\beta}{\alpha}\vec{X}^*, \quad \text{with} \quad \mathbf{A}_{\mathrm{ord}}^T = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ \alpha_{12} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n} & a_{2n} & \ldots & 0 \end{pmatrix}.$$

Since $\mathbf{D}$ is a diagonal matrix and it holds $X_i^* < C_i$, the inverse of $\mathbf{D}$ exists and one may divide this equation by $\mathbf{D}$:

$$\begin{aligned} \frac{\beta}{\alpha}\vec{X}^* &= \mathbf{A}_{\mathrm{ord}}^T \cdot \vec{X}^* \cdot \mathbf{D} \\ \Leftrightarrow \frac{\beta}{\alpha}\vec{X}^* \cdot \mathbf{D}^{-1} &= \mathbf{A}_{\mathrm{ord}}^T \cdot \vec{X}^*, \end{aligned} \tag{A.9}$$

retrieving explicit solutions for the kinase concentrations X_i^* in networks without cycles:

$$\begin{aligned} \frac{\beta}{\alpha} \cdot \frac{X_1^*}{C_1 - X_1^*} &= 0 \cdot X_1^* + 0 \cdot X_2^* + \ldots + 0 \cdot X_n^* = 0 \quad \Rightarrow \quad X_1^* = 0 \\ \frac{\beta}{\alpha} \cdot \frac{X_2^*}{C_2 - X_2^*} &= \alpha_{12} \cdot X_1^* + 0 \cdot X_2^* + \ldots + 0 \cdot X_n^* = 0 \quad \Rightarrow \quad X_2^* = 0 \\ &\vdots \\ \frac{\beta}{\alpha} \cdot \frac{X_n^*}{C_n - X_n^*} &= \alpha_{1n} \cdot X_1^* + \alpha_{2n} \cdot X_2^* + \ldots + 0 \cdot X_n^* = 0 \quad \Rightarrow \quad X_n^* = 0. \end{aligned}$$

$\Rightarrow$For networks without feedback cycles the only steady state solution is the signal off-state: $\vec{X}^* = 0$.

Appendix B

Additional topics to chapter 3

B.1 Summation laws for control of time dependent phenomena

Summation laws in control theory are based on Euler's theorem on homogeneous functions, which can be stated as follows. Let f be a function of $p_1, \ldots, p_n$ such that for all p_i and for all $\lambda > 0$

$$f(\lambda^{\beta_1} p_1, \ldots, \lambda^{\beta_n} p_n) = \lambda^{\gamma} \cdot f(p_1, \ldots, p_n) \qquad \text{for some } \beta_1, \ldots, \beta_n, \gamma \in \mathbf{R}. \tag{B.1}$$

After differentiation of Eq. (B.1) with respect to λ and setting $\lambda = 1$ one ends up with the following equation:

$$\sum_{i=1}^{n} \beta_i \frac{\ln f}{\ln p_i} = \gamma. \tag{B.2}$$

This theorem is used to examine the signalling charateristics, e.g. signalling amplitude, signalling time, signalling duration and integrated response, as homogeneous functions of different variables. For this, one may inspect the system equations that define the time evolution of the signalling components.

The dynamic reaction systems discused here exhibit spatial homogeneity. In those systems, r reactions, numbered with the index i, take place, each at a rate $e_i \cdot v_i$. Through v_i, the rates are functions of concentrations (X_j) of m reactive molecules in the system, and they are each further characterised by an activity e_i. In some systems these e_i correspond with enzyme concentrations, such as the concentration of a protein kinase or a protein phosphatase. A chemical reaction v_i leads to a time-dependent increase in

the concentration of substance X_j:

$$\frac{dX_j}{dt} = \sum_{i=1}^{r} n_{ji} \cdot e_i \cdot v_i(\vec{X}) \tag{B.3}$$

Here n_{ji} is a stoichiometric number which is positive when X_j is a product of the reaction and negative when it is a substrate. The equations define the dynamics of the system in time. We shall assume that the above differential has a unique solution that is asymptotically stable in the sense of Lyapunoff.

Since the reaction rates depend linearly on the enzyme concentrations (activities), simultaneous transformation of these activities and of time

$$\tilde{e}_i = \lambda e_i \qquad , \qquad \tilde{t} = t/\lambda \tag{B.4}$$

leads to a new equation system which coincides with the initial system. Therefore, if the initial conditions are the same, concentrations X_j of the transformed system at the moment t/λ will coincide with concentrations of the initial system at the moment t:

$$X_j(t/\lambda, \lambda\vec{e}) = X_j(t, \vec{e}) = \lambda^0 X_j(t, \vec{e}) \tag{B.5}$$

This shows that the concentrations are homogeneous functions of order zero. By differentiating Eq. (B.5) with respect to the arbitrarily chosen parameter λ one can apply Euler's theorem for homogeneous functions. This leads to the formulation of the summation law for control coefficients of time-dependent concentrations:

$$\sum_{i=1}^{r} C_i^{X_j}(t) - \frac{\partial \ln X_j}{\partial \ln t} = 0. \tag{B.6}$$

At time points where the control by time is zero, e.g. in steady state or at an extreme point in the transient dynamics, the second term is zero. Therefore, the control coefficients on the maximum level of the ERK activation profile (signal amplitude A) sum up to zero:

$$\sum_{i=1}^{r} C_i^{A} = 0. \tag{B.7}$$

The integrated response ("area under the curve") of a component X_j is defined as:

$$I_j(t, e_i) = \int_{\tau=0}^{\tau=t} X_j(\tau, e_i) d\tau. \tag{B.8}$$

Again I_j is considered under transformations of time and enzyme activities given in Eq. (B.5):

$$I_j(t/\lambda, \lambda e_i) = \int_{\tau=0}^{\tau=t/\lambda} X_j(\tau, \lambda e_i) d\tau \overset{\tau=\tau'/\lambda}{\underset{\downarrow}{=}} \int_{\tau'=0}^{\tau'=t} \underbrace{X_j(\tau'/\lambda, \lambda e_i)}_{=X_j(\tau', e_i)} \frac{1}{\lambda} d\tau' = \lambda^{-1} I_j(t, e_i) \tag{B.9}$$

Application of Euler's theorem gives

$$\sum_{i=1}^{r} C_i^{I_j(t)} - C_t^{I_j(t)} = -1. \tag{B.10}$$

If $I_j(t)$ converges for infinite times, this yields the summation law for the integrated response of signalling component X_j:

$$\sum_{i=1}^{r} C_i^{I_j(\infty)} = -1. \tag{B.11}$$

To establish the summation law for the signalling duration two time points $t_{1/2}$ are considered where the concentration X_j attains a certain value $\overline{X}_j$ after pathway stimulation. For example in the calculations in chapter 3.3 $\overline{X}_j$ is chosen as 10% of the signalling amplitude. After integration of Eq. (B.3) these times $t_{1/2}$ are given implicitly by the equations:

$$0 = \int_{\tau=0}^{\tau=t_{1/2}} \sum_{i=1}^{r} n_{ji} \cdot e_i \cdot v_i(\vec{X}) d\tau - \overline{X}_j \quad \Rightarrow \quad t_{1/2} = t_{1/2}(e_i). \tag{B.12}$$

Consider now the transformation where all enzyme activities are activated. The times $t'_{1/2}$ at which the concentration reaches the same magnitude is then found by solving the equation:

$$0 = \int_{\tau=0}^{\tau=t'_{1/2}} \sum_{i=1}^{r} n_{ji} \cdot \lambda \cdot e_i \cdot v_i(\vec{X}) d\tau - \overline{X}_j \overset{\tau=\tau'/\lambda}{\underset{\downarrow}{=}} \int_{\tau'=0}^{\tau'=t'_{1/2}\cdot\lambda} \sum_{i=1}^{r} n_{ji} \cdot e_i \cdot v_i(\vec{X}) d\tau' - \overline{X}_j \tag{B.13}$$

which proves that:

$$\lambda \cdot t'_{1/2} = \lambda \cdot t_{1/2}(\lambda e_i) = t_{1/2}(e_i) \quad \Rightarrow \quad t_{1/2}(\lambda e_i) = \lambda^{-1} t_{1/2}(e_i), \tag{B.14}$$

i.e., times at which a certain concentration is attained are homogeneous functions of the order -1 of the enzyme activities. Therefore, also the signalling duration $d = t_1 - t_2$ is a homogeneous function of the order -1. This gives rise to the summation law for the signalling duration of a component X_j:

$$\sum_{i=1}^{r} C_i^{d_j} = -1. \tag{B.15}$$

Although the summation laws are derived for specific definitions of amplitude, signalling time and duration, they are also valid for other definitions of these quantities.

Interrelations between design principles and dynamic properties of large sets of networks (chapter 2) are most convenient studied by means of Eq. (2.13) and Eq. (2.15). Similar to a time point where a certain concentration $\overline{X}$ is considered, the average signal propagation time τ of a profile $X(t)$ is a homogeneous function of the order -1:

$$\begin{aligned}\tau(\lambda e_i) = \frac{1}{I(\lambda e_i)} \int_0^\infty tX(t, \lambda e_i)\ dt \quad &\overset{t=t'/\lambda}{=} \quad \frac{1}{I(\lambda e_i)} \int_0^\infty t'/\lambda \cdot X(t'/\lambda, \lambda e_i)\ \lambda^{-1} dt' \\ = \frac{1}{\lambda^{-1} I(e_i)} \lambda^{-2} \int_0^\infty t' X(t', e_i)\ dt' \quad &= \quad \lambda^{-1} \cdot \tau(e_i)\ , \end{aligned} \tag{B.16}$$

where the fact is used that for the transformations with respect to λ the functions $X(t, e_i)$ and $I(e_i)$ are homogeneous functions of order zero and -1, respectively. Accordingly, for the signal duration ϑ of $X(t)$ one derives:

$$\begin{aligned}\vartheta(\lambda e_i) \quad &= \quad \sqrt{\frac{1}{I(\lambda e_i)} \int_0^\infty t^2 X_i(t, \lambda e_i)\ dt - \tau(\lambda e_i)^2} \\ &\overset{t=t'/\lambda}{=} \quad \sqrt{\frac{1}{I(\lambda e_i)} \int_0^\infty t'^2 \cdot \lambda^{-2} X_i(t'/\lambda, \lambda e_i)\ \lambda^{-1} dt' - \tau(\lambda e_i)^2} \\ &= \quad \lambda^{-1} \sqrt{\frac{1}{I} \int_0^\infty t'^2 \cdot X_i(t', e_i)\ dt' - \tau(e_i)^2} \\ &= \quad \lambda^{-1} \vartheta(e_i). \end{aligned} \tag{B.17}$$

Applying Euler's theorem gives

$$\sum_{i=1}^{r} C_i^\tau = \sum_{i=1}^{r} C_i^\vartheta = -1. \tag{B.18}$$

The amplitude S of $X(t)$ is defined in Eq. (2.13) as $S = I/2\vartheta$. The logarithmic derivative with respect to reaction rate v_i and the subsequent summation over all processes gives

$$\begin{aligned}\frac{\partial \ln S}{\partial \ln v_i} \quad &= \quad \frac{\partial \ln I}{\partial \ln v_i} - \frac{\partial \ln \vartheta}{\partial \ln v_i} \\ \Rightarrow \quad C_i^S \quad &= \quad C_i^\vartheta - C_i^I \\ \Rightarrow \quad \sum_{i=1}^{r} C_i^S \quad &= \quad \sum_{i=1}^{r} C_i^I - \sum_{i=1}^{r} C_i^\vartheta = -1 - (-1) = 0. \end{aligned} \tag{B.19}$$

B.2 Components, reactions and parameters for the large EGF model

The following Table lists the 103 model components and their inital values (in molecules per cell).

	components	initial concentrations
x1	EGF	0.00000005
x2	EGFR	50000
x3	EGF-EGFR	0
x4	(EGF-EGFR)2	0
x5	(EGF-EGFR*)2	0
x6	EGFRi	0
x7	(EGF-EGFR*)2-GAP-Grb2-Prot	0
x8	(EGF-EGFRi*)2	0
x9	Proti	0
x10	EGF-EGFRi	0
x11	(EGF-EGFRi)2	0
x12	Prot	81000
x13	EGFideg	0
x14	GAP	12000
x15	(EGF-EGFR*)2-GAP	0
x16	EGFi	0
x17	(EGF-EGFRi*)2-GAP	0
x18	(EGF-EGFRi*)2-GAP-Grb2	0
x19	(EGF-EGFRi*)2-GAP-Grb2-Sos	0
x20	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	0
x21	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	0
x22	Grb2	11000
x23	(EGF-EGFR*)2-GAP-Grb2	0
x24	Sos	26300
x25	(EGF-EGFR*)2-GAP-Grb2-Sos	0
x26	Ras-GDP	72000
x27	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP	0
x28	Ras-GTP	0
x29	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP	0
x30	Grb2-Sos	40000
x31	Shc	101000
x32	(EGF-EGFR*)2-GAP-Shc	0
x33	(EGF-EGFR*)2-GAP-Shc*	0
x34	(EGF-EGFR*)2-GAP-Shc*-Grb2	0
x35	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos	0
x36	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP	0
x37	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP	0
x38	Shc*-Grb2-Sos	0

x39	Shc*-Grb2	0
x40	Shc*	0
x41	Raf	40000
x42	Raf-Ras-GTP	0
x43	Ras-GTP*	0
x44	Phosphatase1	40000
x45	Raf*	0
x46	Raf*-phosphatase1	0
x47	MEK	21000000
x48	MEK-Raf*	0
x49	MEK-P	0
x50	MEK-P-Raf*	0
x51	MEK-PP	0
x52	MEK-PP-phosphatase2	0
x53	phosphatase2	40000
x54	MEK-P-phosphatase2	0
x55	ERK	22100000
x56	ERK-MEK-PP	0
x57	ERK-P	0
x58	ERK-P-MEK-PP	0
x59	ERK-PP	0
x60	phosphatase3	10000000
x61	ERK-PP-phosphatase3	0
x62	ERK-P-phosphatase3	0
x63	(EGF-EGFRi*)2-GAP-Shc	0
x64	(EGF-EGFRi*)2-GAP-Shc*	0
x65	(EGF-EGFRi*)2-GAP-Shc*-Grb2	0
x66	(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos	0
x67	(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP	0
x68	(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GTP	0
x69	Ras-GTPi	0
x70	Raf-Ras-GTPi	0
x71	Ras-GTPi*	0
x72	Rafi*	0
x73	Rafi*-phosphatase1	0
x74	MEK-Rafi*	0
x75	MEKi-P	0
x76	MEK-P-Rafi*	0
x77	MEKi-PP	0
x78	MEKi-PP-phosphatase2	0
x79	MEKi-P-phosphatase2	0
x80	ERKi-MEKi-PP	0
x81	ERKi-P	0
x82	ERKi-P-MEKi-PP	0
x83	ERKi-PP	0

x84	ERKi-PP-phosphatase3	0
x85	ERKi-P-phosphatase3	0
x86	EGFRideg	0
x87	(EGF-EGFRi*)2deg	0
x88	(EGF-EGFR*)2-GAP-Grb2-Sos-Prot	0
x89	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP-Prot	0
x90	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP-Prot	0
x91	(EGF-EGFR*)2-GAP-Shc*-Grb2-Prot	0
x92	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Prot	0
x93	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP-Prot	0
x94	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP-Prot	0
x95	(EGF-EGFR*)2-GAP-Grb2-Sos-ERK-PP	0
x96	(EGF-EGFRi*)2-GAP-Grb2-Sos-ERKi-PP	0
x97	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-ERK-PP	0
x98	(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-ERKi-PP	0
x99	(EGF-EGFR*)2-GAP-Grb2-Sos deg	0
x100	(EGF-EGFR*)2-GAP-Grb2-Sos deg	0
x101	Sos-ERK-PP	0
x102	Sos-ERKi-PP	0
x103	Sosi	0

Rate constants of the model are listed in the following Table. First order rate constants are given in $[1/s]$ and second order rate constantsÂ in $[M^{-1}s^{-1}]$.

parameter	value	parameter	value
k1	30000000	kd33	0.2
kd1	0.00384	k33	0.000035
k10	140000	kd34	0.03
k10b	0.054264	kd35	0.0015
kd10b	0	k35	0.0000075
kd10	0.011	k36	0.005
k2	0.0000166	kd36	0
kd2	0.1	kd37	0.3
k3	1	k37	0.0000015
kd3	0.01	k40	0.00005
k4	0.000000173	kd40	0.064
kd4	0.00166	k41	0.00005
kd5	0.0146	kd41	0.0429
k5	0	k42	0.00011833
k6	0.0005	kd42	0.2
kd6	0.005	kd43	1
k8	0.00000166	k43	0
kd8	0.2	kd44	0.01833
k13	42767	kd45	38110
kd13	0	k45	0
k15	10000	kd47	38232

kd15	0	k47	0
k16	0.0000166	k48	0.23833
kd16	0	kd48	0.8
k17	0.0000166	kd49	0.058
kd17	0.06	k49	0
k18	0.000025	k50	0.00000045
kd18	38047	kd50	0.5
k19	0.000000166	kd52	0.033
kd19	0.5	kd53	16
k20	0.0000035	k53	0
kd20	0.4	kd55	38173
k21	0.000000366	k55	0
kd21	0.023	kd56	0.6
k22	0.000035	k56	0.0000235
kd22	0.1	kd57	0.246
k23	6	k57	0
kd23	0.06	k58	0.083333
kd24	0.55	kd58	0.5
k25	0.0000166	k52	0.89072
kd25	0.0214	k44	0.19546
k28	0.00000166	k60	0.0055
kd28	0.0053	kd60	0
k29	0.011667	k61	0.00067
kd29	1	kd61	0
kd32	0.1	kd63	0.275
k32	0.0000004	k63	0
k126	0.000000166	kd126	2
kd127	0.0001	k127	0

The 148 biochemical reactions and the corresponding reaction rates are listed below.

v	reactions	rate equations
1	[EGFR]+[EGF]⟶ [EGF-EGFR]	k1·c1·c2 -kd1·c3
2	[EGF-EGFR]+[EGF-EGFR]⟶[(EGF-EGFR)2]	k2·c3·c3 -kd2·c4
3	[(EGF-EGFR)2]⟶[(EGF-EGFR*)2]	k3·c4·1 -kd3·c5
4	[(EGF-EGFR*)2-GAP-Grb2]+[Prot]⟶[(EGF-EGFR*)2-GAP-Grb2-Prot]	k4·c23·c12 -kd4·c7
5	[(EGF-EGFR*)2-GAP-Grb2-Prot]⟶ [(EGF-EGFRi*)2-GAP-Grb2]+[Proti]	k5·c18·c9 -kd5·c7
6	[EGFR]⟶[EGFRi]	k6·c2·1 -kd6·c6
7	[(EGF-EGFR*)2] ⟶ [(EGF-EGFRi*)2]	k6·c5·1 -kd6·c8
8	[(EGF-EGFR*)2]+[GAP]⟶[(EGF-EGFR*)2-GAP]	k8·c5·c14 -kd8·c15
9	[(EGF-EGFR*)2-GAP]⟶ [(EGF-EGFRi*)2-GAP]	k6·c23·1 -kd6·c18
10	[EGFRi]+[EGFi]⟶[EGF-EGFRi]	k10b·c6·c16 -kd10·c10
11	[EGF-EGFRi]+[EGF-EGFRi]⟶[(EGF-EGFRi)2]	k2·c10·c10 -kd2·c11
12	[(EGF-EGFRi)2]⟶[(EGF-EGFRi*)2]	k3·c11·1 -kd3·c8
13	⟶ [EGFR]	k13·1·1 -kd13·c2
14	[(EGF-EGFRi*)2]+ [GAP]⟶[(EGF-EGFRi*)2-GAP]	k8·c8·c14 -kd8·c17
15	[Proti]⟶ [Prot]	k15·c9·1 -kd15·c12
16	[(EGF-EGFR*)2-GAP]+[Grb2]⟶[(EGF-EGFR*)2-GAP-Grb2]	k16·c22·c15 -kd63·c23

17	[(EGF-EGFR*)2-GAP-Grb2]+[Sos]⟶ [(EGF-EGFR*)2-GAP-Grb2-Sos]	k17·c24·c23 -kd17·c25
18	[(EGF-EGFR*)2-GAP-Grb2-Sos]+[Ras-GDP] ⟶[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP]	k18·c26·c25 -kd18·c27
19	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP]⟶[(EGF-EGFR*)2-GAP-Grb2-Sos]+[Ras-GTP]	k19·c28·c25 -kd19·c27
20	[Ras-GTP*]+[(EGF-EGFR*)2-GAP-Grb2-Sos] ⟶ [(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP]	k20·c25·c43 -kd20·c29
21	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP] ⟶[(EGF-EGFR*)2-GAP-Grb2-Sos]+[Ras-GDP]	k21·c25·c26 -kd21·c29
22	[(EGF-EGFR*)2-GAP]+[Shc]⟶ [(EGF-EGFR*)2-GAP-Shc]	k22·c31·c15 -kd22·c32
23	[(EGF-EGFR*)2-GAP-Shc] ⟶[(EGF-EGFR*)2-GAP-Shc*]	k23·c32·1 -kd23·c33
24	[(EGF-EGFR*)2-GAP-Shc*]+[Grb2]⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2]	k16·c22·c33 -kd24·c34
25	[(EGF-EGFR*)2-GAP-Shc*-Grb2]+[Sos]⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]	k25·c24·c34 -kd25·c35
26	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[Ras-GDP]⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]	k18·c26·c35 -kd18·c36
27	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP] ⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos] + [Ras-GTP]	k19·c35·c28 -kd19·c36
28	[Raf]+[Ras-GTP] ⟶[Raf-Ras-GTP]	k28·c28·c41 -kd28·c42
29	[Raf-Ras-GTP] ⟶[Raf*]+[Ras-GTP*]	k29·c43·c45 -kd29·c42
30	[Ras-GTP*]+[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]	k20·c35·c43 -kd20·c37
31	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP] ⟶ [(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[Ras-GDP]	k21·c35·c26 -kd21·c37
32	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos] ⟶ [(EGF-EGFR*)2-GAP]+[Shc*-Grb2-Sos]	k32·c15·c38 -kd32·c35
33	[Shc*-Grb2-Sos] ⟶ [Grb2-Sos]+[Shc*]	k33·c40·c30 -kd33·c38
34	[(EGF-EGFR*)2-GAP-Grb2-Sos]⟶ [(EGF-EGFR*)2-GAP]+[Grb2-Sos]	k34·c15·c30 -kd34·c25
35	[Grb2-Sos]⟶[Grb2] +[Sos]	k35·c24·c22 -kd35·c30
36	[Shc*]⟶ [Shc]	k36·c40·1 -kd36·c31
37	[(EGF-EGFR*)2-GAP-Shc*]⟶ [(EGF-EGFR*)2-GAP]+[Shc*]	k37·c15·c40 -kd37·c33
38	[Shc*]+[Grb2]⟶ [Shc*-Grb2]	k16·c22·c40 -kd24·c39
39	[(EGF-EGFR*)2-GAP-Shc*-Grb2] ⟶ [(EGF-EGFR*)2-GAP]+[Shc*-Grb2]	k37·c15·c39 -kd37·c34
40	[Shc*-Grb2]+[Sos]⟶[Shc*-Grb2-Sos]	k40·c24·c39 -kd40·c38
41	[(EGF-EGFR*)2-GAP-Shc*] + [Grb2-Sos]⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]	k41·c30·c33 -kd41·c35
42	[Raf*]+[Phosphatase1]⟶ [Raf*-Phosphatase1]	k42·c44·c45 -kd42·c46
43	[Raf*-Phosphatase1] ⟶ [Raf]+[Phosphatase1]	k43·c41·c44 -kd43·c46
44	[MEK] + [Raf*] ⟶[MEK-Raf*]	k44·c47·c45 -kd52·c48
45	[MEK-Raf*] ⟶[MEK-P] +[Raf*]	k45·c49·c45 -kd45·c48
46	[MEK-P]+[Raf*] ⟶[MEK-P-Raf*]	k44·c49·c45 -kd52·c50
47	[MEK-P-Raf*]⟶[MEK-PP] + [Raf*]	k47·c51·c45 -kd47·c50
48	[MEK-PP]+[Phosphatase2]⟶[MEK-PP-Phosphatase2]	k48·c51·c53 -kd48·c52
49	[MEK-PP-Phosphatase2] ⟶[MEK-P] + [Phosphatase2]	k49·c49·c53 -kd49·c52
50	[MEK-P]+[Phosphatase2]⟶ [MEK-P-Phosphatase2]	k50·c53·c49 -kd50·c54
51	[MEK-P-Phosphatase2] ⟶ [MEK]+[Phosphatase2]	k49·c47·c53 -kd49·c54
52	[ERK]+[MEK-PP] ⟶ [ERK-MEK-PP]	k52·c55·c51 -kd44·c56
53	[ERK-MEK-PP]⟶[ERK-P]+[MEK-PP]	k53·c51·c57 -kd53·c56
54	[ERK-P]+[MEK-PP] ⟶ [ERK-P-MEK-PP]	k52·c51·c57 -kd44·c58
55	[ERK-P-MEK-PP] ⟶ [ERK-PP]+[MEK-PP]	k55·c59·c51 -kd55·c58
56	[ERK-PP]+[Phosphatase3]⟶[ERK-PP-Phosphatase3]	k56·c59·c60 -kd56·c61
57	[ERK-PP-Phosphatase3] ⟶[ERK-P]+[Phosphatase3]	k57·c57·c60 -kd57·c61
58	[ERK-P] + [Phosphatase3]⟶ [ERK-P-Phosphatase3]	k58·c60·c57 -kd58·c62
59	[ERK-P-Phosphatase3]⟶ [ERK]+[Phosphatase3]	k57·c55·c60 -kd57·c62
60	[EGFRi]⟶ [EGFRideg]	k60·c6·1 -kd60·c86
61	[EGFi] ⟶ [EGFideg]	k61·c16·1 -kd61·c13
62	[(EGF-EGFRi*)2]⟶[(EGF-EGFRi*)2deg]	k60·c8·1 -kd60·c87
63	[(EGF-EGFRi*)2-GAP]+[Grb2]⟶[(EGF-EGFRi*)2-GAP-Grb2]	k16·c17·c22 -kd63·c18
64	[(EGF-EGFRi*)2-GAP-Grb2]+[Sos]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos]	k17·c24·c18 -kd17·c19
65	[(EGF-EGFRi*)2-GAP-Grb2-Sos]+[Ras-GDP] ⟶[(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP]	k18·c26·c19 -kd18·c20
66	[(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP]⟶[(EGF-EGFRi*)2-GAP-Grb2-Sos]+[Ras-GTPi]	k19·c69·c19 -kd19·c20
67	[Ras-GTPi*]+[(EGF-EGFRi*)2-GAP-Grb2-Sos] ⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTPi]	k20·c71·c19 -kd20·c21
68	[(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP] ⟶[(EGF-EGFRi*)2-GAP-Grb2-Sos]+[Ras-GDP]	k21·c19·c26 -kd21·c21

69	[(EGF-EGFRi*)2-GAP]+[Shc]⟶ [(EGF-EGFRi*)2-GAP-Shc]	k22·c31·c17 -kd22·c63
70	[(EGF-EGFR*)2-GAP-Shc] ⟶[(EGF-EGFR*)2-GAP-Shc*]	k23·c63·1 -kd23·c64
71	[(EGF-EGFRi*)2-GAP-Shc*]+[Grb2]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2]	k16·c22·c64 -kd24·c65
72	[(EGF-EGFRi*)2-GAP-Shc*-Grb2]+[Sos]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]	k25·c24·c65 -kd25·c66
73	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]+[Ras-GDP]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]	k18·c26·c66 -kd18·c67
74	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP] ⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos] + [Ras-GTPi]	k19·c66·c69 -kd19·c67
75	[Raf]+[Ras-GTPi] ⟶[Raf-Ras-GTPi]	k28·c69·c41 -kd28·c70
76	[Raf-Ras-GTPi] ⟶[Rafi*]+[Ras-GTPi*]	k29·c71·c72 -kd29·c70
77	[Ras-GTPi*]+[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]⟶[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]	k20·c71·c66 -kd20·c68
78	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GTP] ⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]+[Ras-GDP]	k21·c66·c26 -kd21·c68
79	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos] ⟶ [(EGF-EGFRi*)2-GAP]+[Shc-Grb2-Sos]	k32·c17·c38 -kd32·c66
80	[(EGF-EGFRi*)2-GAP-Grb2-Sos]⟶ [(EGF-EGFRi*)2-GAP]+[Grb2-Sos]	k34·c17·c30 -kd34·c19
81	[(EGF-EGFRi*)2-GAP-Shc*]⟶ [(EGF-EGFRi*)2-GAP]+[Shc*]	k37·c17·c40 -kd37·c64
82	[(EGF-EGFRi*)2-GAP-Shc*-Grb2] ⟶ [(EGF-EGFRi*)2-GAP]+[Shc*-Grb2]	k37·c17·c39 -kd37·c65
83	[(EGF-EGFRi*)2-GAP-Shc*] + [Grb2-Sos]⟶[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]	k41·c30·c64 -kd41·c66
84	[Rafi*]+[Phosphatase1]⟶ [Rafi*-Phosphatase1]	k42·c44·c72 -kd42·c73
85	[Rafi*-Phosphatase1] ⟶ [Raf]+[Phosphatase1]	k43·c41·c44 -kd43·c73
86	[MEK] + [Rafi*] ⟶[MEK-Rafi*]	k44·c47·c72 -kd52·c74
87	[MEK-Rafi*] ⟶[MEKi-P] +[Rafi*]	k45·c75·c72 -kd45·c74
88	[MEKi-P]+[Rafi*] ⟶[MEK-P-Rafi*]	k44·c72·c75 -kd52·c76
89	[MEK-P-Rafi*]⟶[MEKi-PP] + [Rafi*]	k47·c72·c77 -kd47·c76
90	[MEKi-PP]+[Phosphatase2]⟶[MEKi-PP-Phosphatase2]	k48·c77·c53 -kd48·c78
91	[MEKi-PP-Phosphatase2] ⟶[MEKi-P] + [Phosphatase2]	k49·c75·c53 -kd49·c78
92	[MEKi-P]+[Phosphatase2]⟶ [MEKi-P-Phosphatase2]	k50·c53·c75 -kd50·c79
93	[MEKi-P-Phosphatase2] ⟶ [MEK]+[Phosphatase2]	k49·c47·c53 -kd49·c79
94	[ERK]+[MEKi-PP] ⟶ [ERK-MEKi-PP]	k52·c55·c77 -kd44·c80
95	[ERK-MEKi-PP]⟶[ERKi-P]+[MEKi-PP]	k53·c81·c77 -kd53·c80
96	[ERKi-P]+[MEKi-PP] ⟶ [ERKi-P-MEKi-PP]	k52·c77·c81 -kd44·c82
97	[ERKi-P-MEKi-PP] ⟶ [ERKi-PP]+[MEKi-PP]	k55·c83·c77 -kd55·c82
98	[ERKi-PP]+[Phosphatase3]⟶[ERKi-PP-Phosphatase3]	k56·c83·c60 -kd56·c84
99	[ERKi-PP-Phosphatase3] ⟶[ERKi-P]+[Phosphatase3]	k57·c81·c60 -kd57·c84
100	[ERKi-P] + [Phosphatase3]⟶ [ERKi-P-Phosphatase3]	k58·c60·c81 -kd58·c85
101	[ERKi-P-Phosphatase3]⟶ [ERK]+[Phosphatase3]	k57·c55·c60 -kd57·c85
102	[(EGF-EGFR*)2-GAP]⟶ [(EGF-EGFRi*)2-GAP]	k6·c15·1 -kd6·c17
103	[(EGF-EGFR*)2-GAP-Shc]⟶ [(EGF-EGFRi*)2-GAP-Shc]	k6·c32·1 -kd6·c63
104	[(EGF-EGFR*)2-GAP-Shc*]⟶ [(EGF-EGFRi*)2-GAP-Shc*]	k6·c33·1 -kd6·c64
105	[(EGF-EGFR*)2-GAP-Grb2-Sos]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos]	k6·c25·1 -kd6·c19
106	[(EGF-EGFR*)2-GAP-Grb2-Sos]+[Prot]⟶[(EGF-EGFR*)2-GAP-Grb2-Sos-Prot]	k4·c25·c12 -kd4·c88
107	[(EGF-EGFR*)2-GAP-Grb2-Sos-Prot]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos]+[Proti]	k5·c9·c19 -kd5·c88
108	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP]	k6·c27·1 -kd6·c20
109	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP]+[Prot]⟶[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP-Prot]	k4·c27·c12 -kd4·c89
110	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP-Prot]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP]+[Proti]	k5·c9·c20 -kd5·c89
111	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP]	k6·c29·1 -kd6·c21
112	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP]+[Prot]⟶[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP-Prot]	k4·c29·c12 -kd4·c90
113	[(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP-Prot]⟶ [(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP]+[Proti]	k5·c9·c21 -kd5·c90
114	[(EGF-EGFR*)2-GAP-Shc*-Grb2]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2]	k6·c34·1 -kd6·c65
115	[(EGF-EGFR*)2-GAP-Shc*-Grb2]+[Prot]⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Prot]	k4·c34·c12 -kd4·c91
116	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Prot]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2]+[Proti]	k5·c9·c65 -kd5·c91
117	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]	k6·c35·1 -kd6·c66
118	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[Prot]⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Prot]	k4·c35·c12 -kd4·c92
119	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Prot]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]+[Proti]	k5·c9·c66 -kd5·c92

120	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]	k6·c36·1 -kd6·c67
121	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]+[Prot]⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP-Prot]	k4·c36·c12 -kd4·c93
122	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GDP-Prot]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]+[Proti]	k5·c9·c67 -kd5·c93
123	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]	k6·c37·1 -kd6·c68
124	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]+[Prot]⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP-Prot]	k4·c37·c12 -kd4·c94
125	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-Ras-GTP-Prot]⟶ [(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]+[Proti]	k5·c68·c9 -kd5·c94
126	[(EGF-EGFR*)2-GAP-Grb2-Sos]+[ERK-PP] ⟶[(EGF-EGFR*)2-GAP-Grb2-Sos-ERK-PP]	k126·c59·c25 -kd126·c95
127	[(EGF-EGFRi*)2-GAP-Grb2-Sos]+[ERKi-PP] ⟶[(EGF-EGFRi*)2-GAP-Grb2-Sos-ERKi-PP]	k126·c83·c19 -kd126·c96
128	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]+[ERK-PP] ⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-ERK-PP]	k126·c59·c35 -kd126·c97
129	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]+[ERKi-PP] ⟶[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-ERKi-PP]	k126·c83·c66 -kd126·c98
130	[Sos]+[ERK-PP]⟶[Sos-ERK-PP]	k126·c59·c24 -kd126·c101
131	[Sos]+[ERKi-PP]⟶[Sos-ERKi-PP]	k126·c83·c24 -kd126·c102
132	[(EGF-EGFRi*)2-GAP]⟶[(EGF-EGFRi*)2deg]	k60·c17·1 -kd60·c87
133	[(EGF-EGFRi*)2-GAP-Grb2]⟶[(EGF-EGFRi*)2deg]	k60·c18·1 -kd60·c87
134	[(EGF-EGFRi*)2-GAP-Grb2-Sos]⟶[(EGF-EGFRi*)2deg]	k60·c19·1 -kd60·c87
135	[(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP]⟶[(EGF-EGFRi*)2deg]	k60·c20·1 -kd60·c87
136	[(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP]⟶[(EGF-EGFRi*)2deg]	k60·c21·1 -kd60·c87
137	[(EGF-EGFRi*)2-GAP-Shc]⟶[(EGF-EGFRi*)2deg]	k60·c63·1 -kd60·c87
138	[(EGF-EGFRi*)2-GAP-Shc*]⟶[(EGF-EGFRi*)2deg]	k60·c64·1 -kd60·c87
139	[(EGF-EGFRi*)2-GAP-Shc*-Grb2]⟶[(EGF-EGFRi*)2deg]	k60·c65·1 -kd60·c87
140	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]⟶[(EGF-EGFRi*)2deg]	k60·c66·1 -kd60·c87
141	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GDP]⟶[(EGF-EGFRi*)2deg]	k60·c67·1 -kd60·c87
142	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-Ras-GTP]⟶[(EGF-EGFRi*)2deg]	k60·c68·1 -kd60·c87
143	[(EGF-EGFR*)2-GAP-Grb2-Sos-ERK-PP] ⟶[(EGF-EGFR*)2-GAP-Grb2-Sos]deg+[ERK-PP]	k127·c59·c99 -kd127·c95
144	[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos-ERK-PP] ⟶[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos]deg+[ERK-PP]	k127·c59·c99 -kd127·c97
145	[Sos-ERK-PP]⟶[Sosi]+[ERK-PP]	k127·c59·c103 -kd127·c101
146	[(EGF-EGFRi*)2-GAP-Grb2-Sos-ERKi-PP] ⟶[(EGF-EGFRi*)2-GAP-Grb2-Sos]deg+[ERKi-PP]	k127·c83·c100 -kd127·c96
147	[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos-ERKi-PP] ⟶[(EGF-EGFRi*)2-GAP-Shc*-Grb2-Sos]deg+[ERKi-PP]	k127·c83·c100 -kd127·c98
148	[Sos-ERK-PPi]⟶[Sosi]+[ERK-PPi]	k127·c83·c103 -kd127·c102

B.3 Localisation of control coefficients for the signal amplitude, signal duration and integrated response of the ERK-PP activation profile

According to Table 3.1 eight most substantial control coefficients (for small perturbation) on the amplitude of the ERK-PP activation profile are found. The locations of these (five positive and three negative) coefficients within the EGF signalling network are depicted in Fig. B.1. The positive and negative control coefficients are displayed in red and blue, respectively.

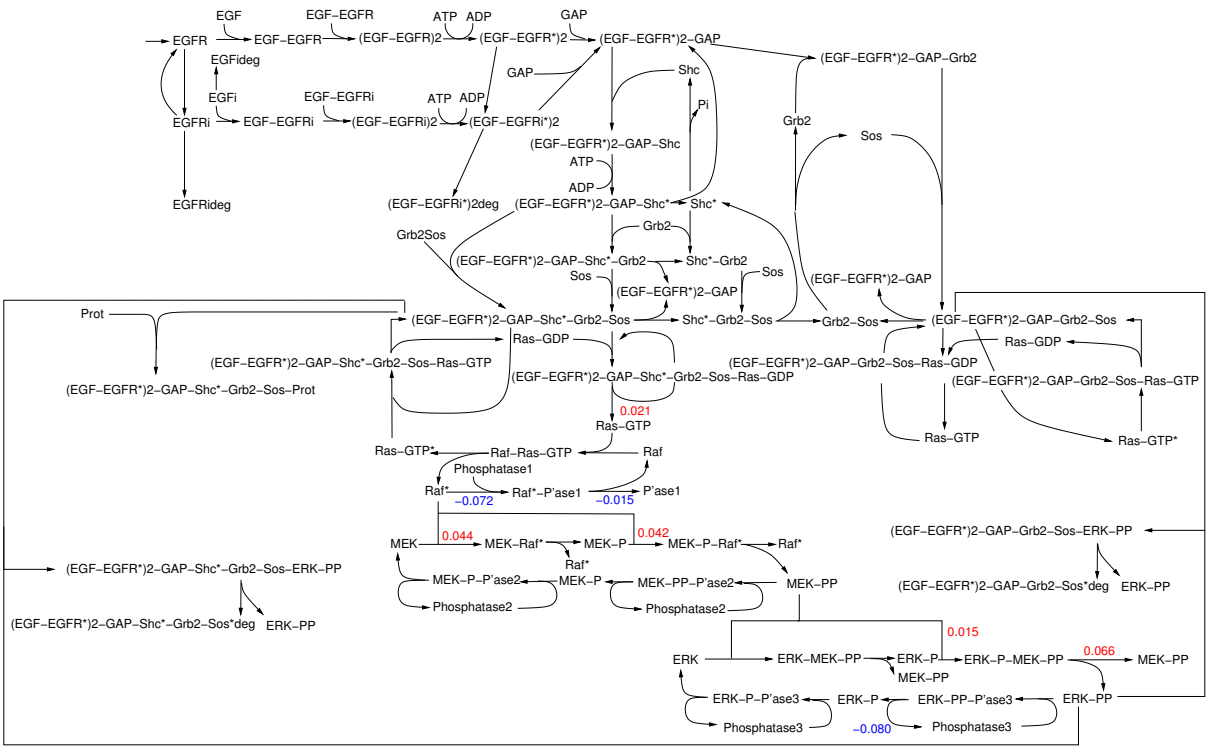


Figure B.1: Localisation of most significant control coefficients for the signal amplitude.

There are 18 (nine positive and nine negative) most substantial control coefficients for the duration of the ERK-PP activation profile. How they are distributed in the network is shown in Fig. B.2.

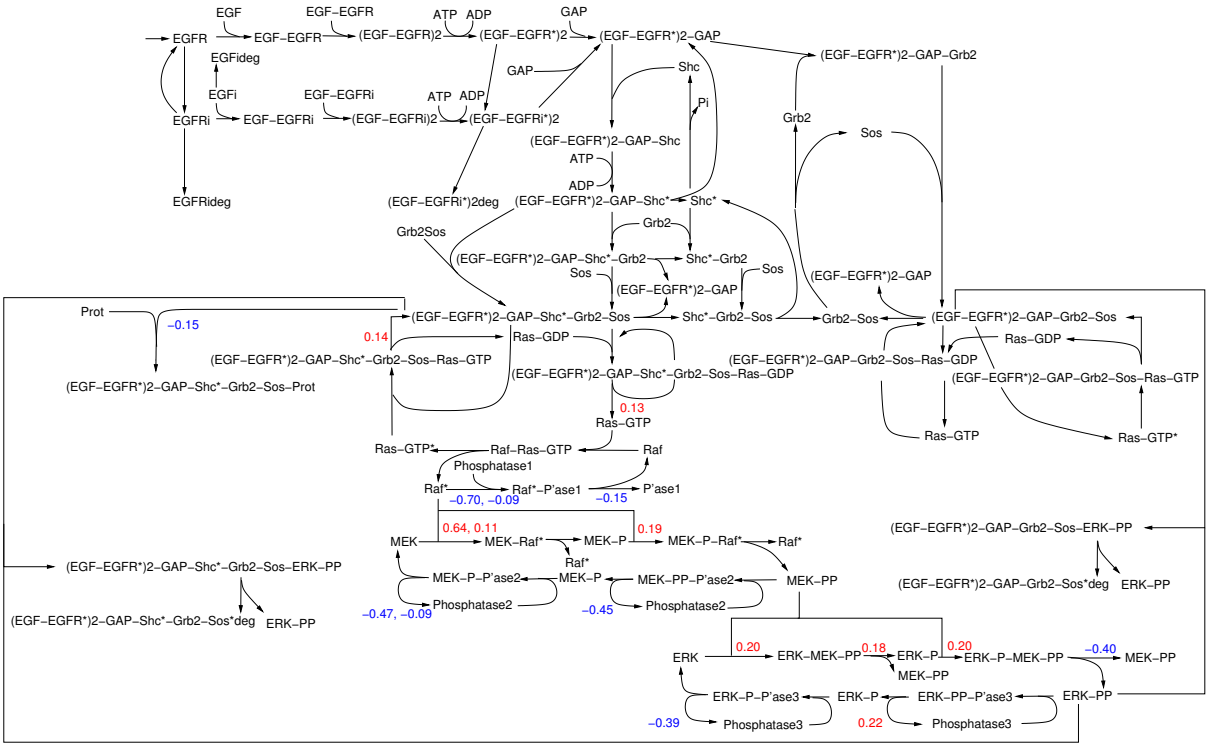


Figure B.2: Localisation of most significant control coefficients for signal duration.

Figure B.3 depicts how the most substantial control coefficients of the integrated response (eight positive and seven negative) are arranged in the network.

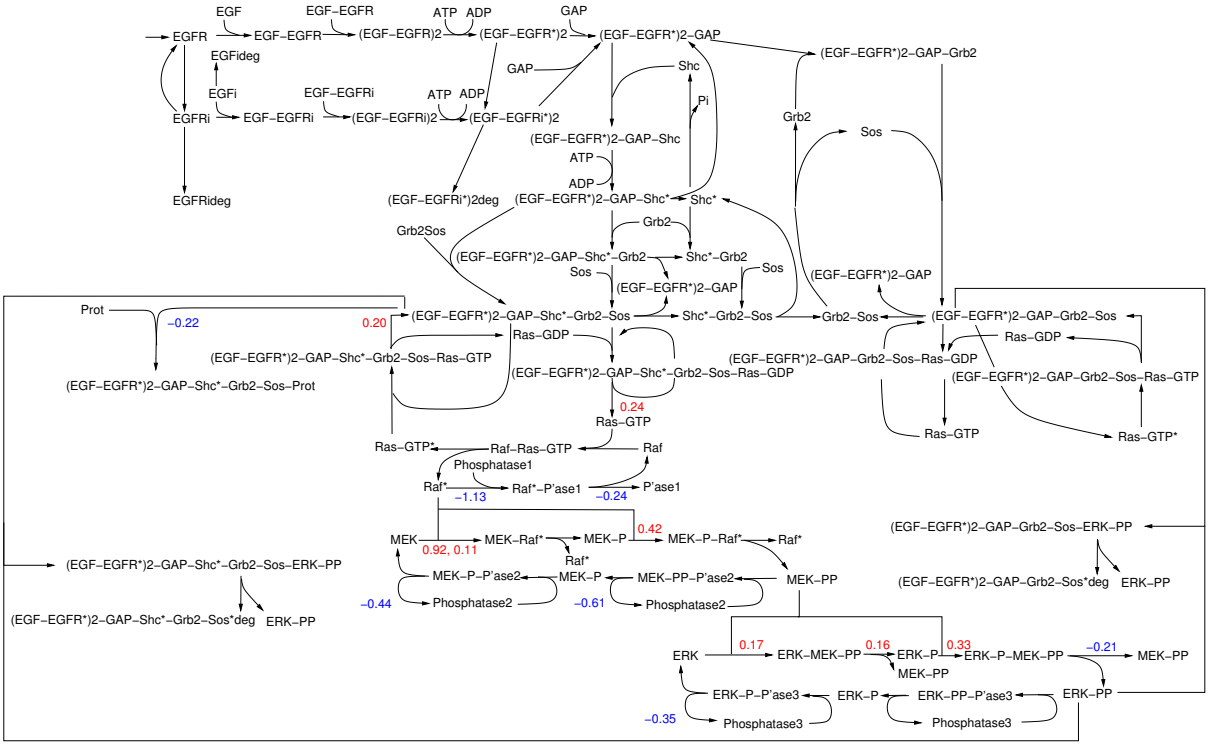


Figure B.3: Localisation of most significant control coefficients for integrated response.

B.4 Kinase cascade model for a direct comparison with inhibitor experiments

The model describes a linear signal transduction cascade containing the following eight irreversible reactions. Reaction velocities are given by Michaelis-Menten kinetics. The model is also available on the website: http://www.bio.vu.nl/hwconf/supplements/. The reaction rates with the initial set of parameters used for each reaction are given in Table B.4.

process	rate equation	initial concentration
$R \rightarrow R_i$	$v_1 = \frac{R}{0.1 + R}$	$R = 0.5$
$R_i \rightarrow R$	$v_2 = \frac{0.01 \cdot R_i}{0.1 + R_i}$	$R_i = 0$
$X_1 \rightarrow X_1^*$	$v_3 = \frac{R \cdot X_1}{0.1 + X_1}$	$X_1 = 1$
$X_1^* \rightarrow X_1$	$v_4 = \frac{0.3 \cdot X_1^*}{1 + X_1^*}$	$X_1^* = 0$
$X_2 \rightarrow X_2^*$	$v_5 = \frac{X_1^* \cdot X_2}{0.1 + X_2}$	$X_2 = 1$
$X_2^* \rightarrow X_2$	$v_6 = \frac{0.3 \cdot X_2^*}{0.1 + X_2^*}$	$X_2^* = 0$
$X_3 \rightarrow X_3^*$	$v_7 = \frac{X_2^* \cdot X_3}{0.1 + X_3}$	$X_3 = 1$
$X_3^* \rightarrow X_3$	$v_8 = \frac{0.3 \cdot X_3^*}{1 + X_3^* + Y}$	$X_3^* = 0$, $Y = 0$

Table B.4: Reaction rates and initial set of parameters for the kinase cascade model including receptor deactivation, receptor recycling, kinase and phosphatase reactions. The initial concentrations at $t = 0$ are also given.

Y represents the concentration of a competitive inhibitor of the phosphatase of X_3^*. The concentration Y is normally set to 0. In the case where the effect of inhibition of

the third phosphatase is demonstrated, Y is set from 0 to 2, 4 and 6, respectively (cf. Figure 3.3B). The effect of a non-competitive kinase inhibitor is studied by decreasing the $V_{\max}$ of v_5 from 1 to 0.8, 0.6 and 0.4, respectively (cf. Figure 3.3A).

The signal is taken to reside in the extent of phosphorylation of the third kinase in the model MAPK pathway (Figure 3.2). The characteristics of that signal are its maximum (amplitude), its ultimate value (final strength), its duration and its time-integrated concentration. Control coefficients are calculated by increasing the rate of a reaction by 1% and then divide the relative effect on the characteristics by the relative change in reaction rate.

For the initial set of parameters (Figure 3.3) the control coefficients are given in Table B.5.

reaction	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	$\sum$
amplitude	-1.20	0.03	1.21	-0.81	1.15	-0.77	1.06	-0.64	0.02
duration	-0.44	0.17	0.44	-0.60	0.42	-0.62	0.38	-0.70	-0.96
int. response	-1.41	0.12	1.43	-1.25	1.35	-1.24	1.22	-1.16	-0.94
final strength	-1.03	1.04	1.03	-1.02	1.03	-1.02	1.02	-1.02	0.04

Table B.5: Control of any out of four characteristics of the time dependent signal, by any of the processes in the cascade. The kinases have control coefficients on the X_3^* amplitude which are about three times higher than their coefficients on X_3^* duration. Their control coefficients on duration are moreover smaller (in absolute values) than the duration coefficients from the phosphatases.

B.5 Experimental setup for inhibitor experiments and western blot analysis

For the cell culture NRK fibroblasts are cultured in Dulbecco's modified Eagle's medium (Biowhittaker Europe), supplemented with 10% fetal bovine serum (Gibco), 0.10 g/l penicillin and 0.10 g/l streptomycin in a humidified 5% CO_2 incubator at 37 degree celsius. For serum-starvation, cells are washed then with 1x Hank's buffered salt solution (Gibco) and then used the same medium, but with 0.5% bovine serum albumin (Applichem) instead of serum.

For the stimulation experiments the cells are grown in culture dishes to sub-confluency and then serum-starved for three days in order to be arrested in the G_0-phase of the cell cycle. Subsequently cells are stimulated with 10 ng/ml EGF (Becton Dickinson) for the indicated periods of time. Where indicated MEK is inhibited by preincubation for 1 hour with various concentrations of the non-competitive inhibitor PD98059 (Alessi et al. [1995]). For the phosphatase inhibition (PTPs) cells are pre-incubated for 1 hour with 0.20 mM sodium orthovanadate (Huyer et al. [1997]).

Western blot analysis is used to quantify the time profiles of the kinase activations.

After stimulation, cells are washed twice with ice-cold phosphate-buffered saline (17 mM NaH_2PO_4, 38.5 mM NaH_2PO_4, 68 mM NaCl, pH 7.4) and incubated on ice with 'lysis buffer' (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% SDS, 0.1% NonidetP40, 0.1% sodium deoxycholate, 50 mM NaF, 1 mM sodium-orthovanadate, 1x Complete protease inhibitor mix (Roche)) for 20 minutes. Cell lysates are scraped using a cell scraper (25cm/1.8cm, Costar), collected, vortexed for 10 seconds, frozen in liquid nitrogen and stored at -80 degree celsius. Protein contents in the cell lysates are determined with the BCA assay (Pierce). Proteins are separated by SDS-polyacrylamide gel electrophoresis, using a 12% resolving gel. Exactly 10μg of total protein of each sample is loaded onto the gel in 'loading buffer' (0.25 M Tris (HCl) pH 7.6, 8% SDS, 40% glycerol, 0.05% bromophenol blue, 20 mM dithiothreitol). Proteins are electro-transferred to Immun-BlotTM polyvinylidene difluoride membranes (Bio-Rad) using a current of $0.40A$ overnight at 4 degree celsius. Membranes are washed in Tris-buffered saline (TBS: 20 mM Tris-HCl, pH 7.6, 150 mM NaCl) supplemented with 0.05% Tween 80 (TBS-T), pre-incubated for 1 hour at room temperature with blocking buffer (5% skim milk powder (Oxoid) in TBS-T), supplemented with 0.5 mM Na_3VO_4, and incubated overnight at 4 degree celsius with anti-phospho-p44/42 MAP kinase monoclonal antibody (Cell Signalling) in blocking buffer (1:2000), supplemented with 0.5 mM Na_3VO_4. After washing, membranes are incubated for 1 hour at room temperature with horse-radish peroxidase-conjugated goat anti-mouse IgG (Bio-Rad) in blocking buffer (1:3000). Membranes are washed again and then incubated for 5 minutes with Lumi-Light$^{\mathrm{PLUS}}$ Western Blotting Substrate (Roche). Signals are detected with a FluorS$^{\mathrm{TM}}$ MultiImager (Bio-Rad) and quantified using the Multi-Analist software (Bio-Rad).

Bibliography

N. G. Ahn and E. G. Krebs. Evidence for an epidermal growth factor-stimulated protein kinase cascade in swiss 3T3 cells. Activation of serine peptide kinase activity by myelin basic protein kinases in vitro. *J. Biol. Chem.*, 265:11495–11501., 1990.

N. G. Ahn, R. Seger, R. L. Bratlien, C. D. Diltz, N. K. Tonks, and E. G. Krebs. Multiple components in an epidermal growth factor-stimulated protein kinase cascade. In vitro activation of a myelin basic protein/microtubule-associated protein 2 kinase. *J. Biol. Chem.*, 266:4220–4227, 1991.

D. R. Alessi, A. Cuenda, P. Cohen, D. T. Dudley, and A. R. Saltiel. Pd098059 is a specific inhibitor of the activation of mitogen-activated protein kinase kinase in vitro and in vivo. *J. Biol. Chem.*, 270:27489–27494, 1995.

A. R. Asthagiri and D. A. Lauffenburger. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol. Progr.*, 17:227–239, 2001.

A. R. Asthagiri, C. A. Reinhart, A. F Horwitz, and D. A. Lauffenburger. The role of transient ERK2 signals in fibronectin- and insulin-mediated DNA synthesis. *J. Cell Sci.*, 113:4499–4510, 2000.

U. S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–387, 1999.

U. S. Bhalla and R. Iyengar. Robustness of the bistable behavior of a biological signaling feedback loop. *Chaos*, 11(1):221–226, 2001.

U. S. Bhalla, P. T. Ram, and R. Iyengar. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297:1018–1023, 2002.

F. A. Brightman and D. A. Fell. Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. *FEBS Lett.*, 482:169–174, 2000.

W. R Burack and T. W. Sturgill. The activating dual phosphorylation of MAPK by MEK is nonprocessive. *Biochemistry*, 36:5929–5933, 1997.

K. Y. Choi, B. Satterberg, D. M. Lyons, and E. A. Elion. Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in S. cerevisiae. *Cell*, 78: 499–512, 1994.

M. H. Cobb. MAP kinase pathways. *Progr. Biophys. Bio.*, 71:479–500, 1999.

S. J. Cook, N. Aziz, and M. McMahon. The repertoire of fos and jun proteins expressed during the G1 phase of the cell cycle is determined by the duration of mitogen-activated protein kinase activation. *Mol. Cell. Biol.*, 19:330–341, 1999.

K. J. Cowan and K. B. Storey. Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress. *J. Exp. Biol.*, 206: 1107–1115, 2003.

O Ebenhöh and R. Heinrich. Stoichiometric design of metabolic networks: Multifunctionality, clusters, optimization, weak and strong robustness. *Bull. Math. Biol.*, 65: 323–357, 2003.

E. M. Ferrell, J. E. and Machhleder. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science*, 280:895–898, 1998.

J. E. Ferrell and R. R. Bhatt. Mechanistic studies of the dual phosphorylation of mitogen-activated protein kinase. *J. Biol. Chem.*, 272(30):19008–19016, 1997.

A. Goldbeter and D. E. Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci.*, 78:6840–6844, 1981.

A. Goldbeter and D. E. Koshland. Sensitivity amplification in biochemical systems. *Rev. Biophys. Q.*, 15:555–591, 1982.

N. R. Gough. Science's signal transduction knowledge environment: the connections maps database. *Ann N Y Acad Sci.*, 971:585–587, 2002.

F. Harary. *Graph Theory. Polya's Enumeration Theorem.* Reading, MA: Addison-Wesley, 1994.

R. Heinrich and T. Rapoport. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.*, 42:89–95, 1974.

R. Heinrich, T. Rapoport, and G. Neel. Mathematical models of protein kinase signal transduction. *Mol. Cell*, 9:957–970, 2002.

R. Heinrich, S. Schuster, and H. G. Holzhütter. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur. J. Biochem.*, 201:1–21, 1991.

R. Hoshino, Y. Chatani, T. Yamori, T. Tsuruo, H. Oka, O. Yoshida, Y. Shimada, S. Ari-i, H. Wada, J. Fujimoto, and M. Kohno. Constitutive activation of the 41-/43-kDa mitogen-activated protein kinase signaling pathway in human tumors. *Oncogene*, 18: 813–822, 1999.

C. Y. Huang and J. E. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. USA*, 93:10078–10083, 1996.

G. Huyer, S. Liu, J. Kelly, J. Moffat, P. Payette, B. Kennedy, G. Tsaprailis, M. J. Gresser, and C. Ramachandran. Mechanism of inhibition of protein-tyrosine phosphatases by vanadate and pervanadate. *J. Biol. Chem.*, 272:843–851, 1997.

H. Kacser and J. A. Burns. The control of flux. *Symp. Soc. Exp. Biol.*, 27:65–104, 1973.

B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, 267:1583–1588, 2000.

B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.*, 274:30169–30181, 1999.

B. N. Kholodenko, J. B. Hoek, H. V. Westerhoff, and G. C. Brown. Quantification of information transfer via cellular signal transduction pathways. *FEBS Letters*, 414: 430–434, 1997.

J. M. Kyriakis, H. App, X. F. Zhang, P. Banerjee, D. L. Brautigan, U. R. Rapp, and J. Avruch. Raf-1 activates MAP kinase-kinase. *Nature*, 358:417–421, 1992.

J. M. Kyriakis and J. Avruch. Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiological Reviews*, 81(2):807–869, 2001.

D. H Lahaye, M. G Camps, P. E. Erp, P. H. Peters, and E. J. Zoelen. Epidermal growth factor (EGF) receptor density controls mitogenic activation of normal rat kidney (NRK) cells by EGF. *J. Cell Physiol.*, 174:9–17, 1998.

E. Lee, A. Salic, R. Krüger, R. Heinrich, and M. W. Kirschner. The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS Biol.*, 1(1):116–132, 2003.

A. Levchenko, J. Bruck, and P. W. Sternberg. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. USA*, 97:5818–5823, 2000.

G. Manning, G. D. Plowman, T. Hunter, and S. Sudarsanam. Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences*, 27:514–520, 2002.

C. J. Marshall. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell*, 80:179–185, 1995.

W. T. Matten, T. D. Copeland, N. G. Ahn, and G. F. Vande Woude. Positive feedback between MAP kinase and Mos during xenopus oocyte maturation. *Dev. Biol.*, 179: 485–492, 1996.

E. Meléndez-Hevia, T. G. Waddell, R. Heinrich, and F. Montero. Theoretical approaches to the evolutionary optimization of glycolysis; chemical analysis. *Eur. J. Biochem.*, 244:527–543, 1997.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.

G. Pagès, P. Lenormand, G. L'Allemain, J. C. Chambard, S. Meloche, and J. Pouyssegur. Mitogen-activated protein kinases p42mapk and p44mapk are required for fibroblast proliferation. *Proc. Natl. Acad. Sci. USA*, 90:8319–8323, 1993.

J. Saez-Rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, and E. D. Gilles. Modular analysis of signal transduction networks. *Control Systems Magazine, IEEE*, 24(4):35–52, 2004.

F. Schacherer, C. Choi, U. Götze, M. Krull, S. Pistor, and E. Wingender. The transpath signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*, 17:1053–1057, 2001.

B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Muller. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnology*, 20(4):370–375, 2002.

J. S. Sebolt-Leopold. Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene*, 19:6594–6599, 2000.

O. J. Somsen, M. Siderius, F. F. Bauer, J. L. Snoep, and H. V. Westerhoff. Selectivity in overlapping MAP kinase cascades. *J. Theor. Biol.*, 218:343–354, 2002.

A. Stephani, J. C. Nuno, and R. Heinrich. Optimal stoichiometric design of ATP-producing systems as determined by an evolutionary algorithm. *J. Theor. Biol.*, 199: 45–61, 1999.

P. S. Swain and E. D. Siggia. The role of proofreading in signal transduction specificity. *Biophys. J.*, 82:2928–2933, 2002.

R. M. Tombes, K. L. Auer, R. Mikkelsen, K. Valerie, M. P. Wymann, C. J. Marshall, M. McMahon, and P. Dent. The mitogen-activated protein (MAP) kinase cascade can either stimulate or inhibit DNA synthesis in primary cultures of rat hepatocytes depending upon whether its activation is acute/phasic or chronic. *Biochem. J.*, 330: 1451–1460, 1998.

E. J. van Zoelen. Phenotypic transformation of normal rat kidney cells: a model for studying cellular alterations in oncogenesis. *Crit. Rev. Oncog.*, 2:311–333, 1991.

K Yoshioka. Scaffold proteins in mammalian MAP kinase cascades. *J. Biochem.*, 135: 657–661, 2004.

Lebenslauf

Persönliche Daten	Bernd Binder Tempelhofer Ufer 12 10963 Berlin Tel. 030-25899524 E-mail: bernd.binder@rz.hu-berlin.de
Schulausbildung	Grundschule in Vellberg und Gymnasium bei St. Michael in Schwäbisch Hall, Abitur Mai 1990
Universitätsausbildung	Physikstudium an der Universität Fridericiana Karlsruhe, WS 1991 / 92 – SS 1994 Physikstudium an der Ruprecht-Karls-Universität Heidelberg, SS 1994 – SS 1998
Hochschulabschluss	Diplom Physik, September 1998
Berufserfahrung	Wissenschaftlicher Mitarbeiter am Deutschen Krebsforschungszentrum (DKFZ) in Heidelberg, Arbeitsgruppe theoretische Bioinformatik Wissenschaftlicher Mitarbeiter am Max Planck Institut für molekulare Genetik in Berlin, Arbeitsgruppe Bioinformatik Wissenschaftlicher Mitarbeiter an der Humboldt-Universität zu Berlin, Institut für Biologie, Arbeitsgruppe theoretische Biophysik, seit November 2001
Beginn der Promotion	1. November 2001
Auslandsaufenthalte	Forschungsaufenthalt an der Vrije Universiteit Amsterdam, Niederlande, Okt. - Dez. 2003
Fremdsprachen	Englisch, sehr gute Kenntnisse in Wort und Schrift Französisch, fortgeschrittene Kenntnisse Türkisch, Konversation

Berlin, den 31. Januar 2005

Publikationsliste

Binder, B. and Heinrich, R. (2002). Dynamic stability of signal transduction networks depending on downstream and upstream specificity of protein kinases. *Mol. Biol. Rep. 29*, 51-55.

Binder, B. and Heinrich, R. (2004). Interrelations between dynamical properties and structural characteristics of signal transduction networks. *Genome Informatics 15*, 13-23.

Hornberg, J. and Bruggeman, F. and Binder, B. and Geest, C. and Marjolein Bij de Vaate, A. and Lankelma, J. and Heinrich, R. and Westerhoff, H. Principles behind the Multifarious Control of Signal Transduction: ERK phosphorylation and kinase/phosphatase control. (2005) *FEBS J. 272*, 244-258.

Hornberg, J. and Binder, B. and Bruggeman, F. and Schoeberl, B. and Heinrich, R. and Westerhoff, H. (2004). Control of MAPK signalling: from complexity to what really matters. Submitted to *Oncogene*

Danksagung

An erster Stelle möchte ich Herrn Prof. R. Heinrich für seine Betreuung und das Interesse, das er meiner Arbeit entgegengebracht hat, danken. Besonders bedanken möchte ich mich auch dafür, dass er es mir ermöglichte, an den vielen Konferenzen in verschiedenen Ländern teilzunehmen.
Herrn Prof. Hans Peter Herzel danke ich für seine hilfreichen Hinweise und Verbesserungsvorschläge. Ihm und Herrn Prof. Thomas Höfer danke ich für das Interesse, das sie meiner Arbeit entgegengebracht haben.
Ein wesentlicher Faktor für das Gelingen dieser Arbeit war die konstruktive und angenehme Arbeitsatmosphäre in der Arbeitsgruppe. Dafür möchte ich allen Mitgliedern - auch den ehemaligen - meinen Dank aussprechen.
Besonders hervorheben möchte ich meinen Dank an Jana Schütze und Oliver Ebenhöh für ihre Bereitschaft, meine Arbeit korrekturzulesen.

Erklärung

Ich versichere hiermit, die vorliegende Arbeit selbständig und ausschließlich unter Verwendung der angegebenen Mittel und ohne unerlaubte Hilfen angefertigt zu haben.

Berlin, den 31. Januar 2005

Bernd Binder